

AD-A173 411

ACOUSTO-OPTIC PROCESSING OF 2-D SIGNALS USING TEMPORAL
AND SPATIAL INTEGR. (U) CALIFORNIA INST OF TECH
PASADENA DEPT OF ELECTRICAL ENGINEER. D PSALTIS

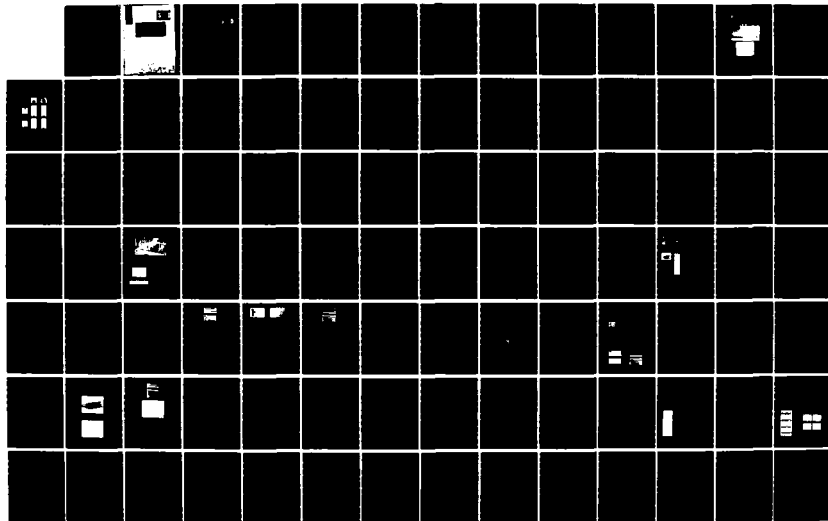
1/2

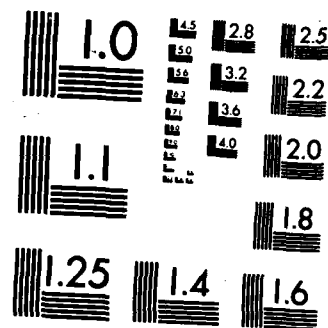
UNCLASSIFIED

29 APR 86 AFOSR-TR-86-0847 AFOSR-82-0120

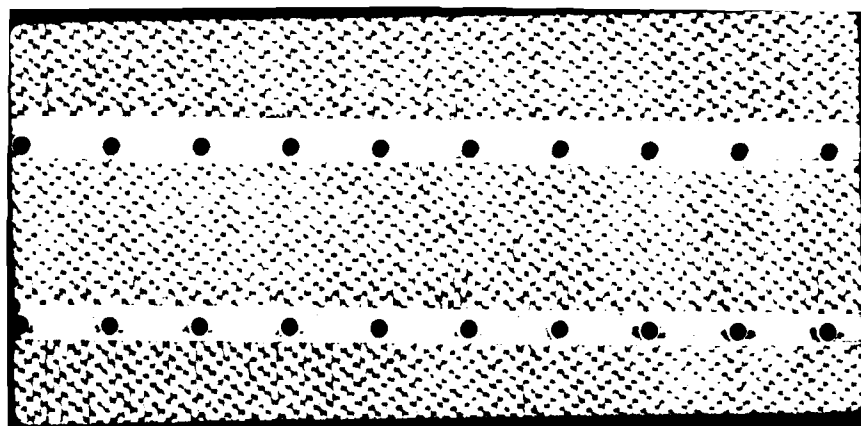
F/G 17/9

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



DTIC
ELECTE
OCT 21 1986
S D D

FINAL REPORT

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Grant Number AFOSR-82-0128

**ACOUSTO-OPTIC PROCESSING OF 2-D SIGNALS
USING TEMPORAL AND SPATIAL INTEGRATION**

**Principal Investigator:
Professor Demetri Psaltis
California Institute of Technology
Pasadena, CA, 91125**

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION		1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION/AVAILABILITY OF REPORT	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE		Approved for public release; distribution unlimited.	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AFOSR-82-0128		5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 86-0847	
6a. NAME OF PERFORMING ORGANIZATION California Institute of Technology		7a. NAME OF MONITORING ORGANIZATION AFOSR	
6b. OFFICE SYMBOL (If applicable)		7b. ADDRESS (City, State and ZIP Code) Same as 8c	
6c. ADDRESS (City, State and ZIP Code) Department of Electrical Engineering MS 116-81 Pasadena, CA 91125		8. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR- 82-0128	
9a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR		9b. OFFICE SYMBOL (If applicable) nm	
10a. ADDRESS (City, State and ZIP Code) AFOSR Bldg 410 BAF B3 DC 20332		10. SOURCE OF FUNDING NOS.	
11. TITLE (Include Security Classification) OVER		PROGRAM ELEMENT NO. 61102F	
12. PERSONAL AUTHOR(S) Demetri Psaltis		PROJECT NO. 2305	
13a. TYPE OF REPORT Final		TASK NO. B1	
13b. TIME COVERED FROM _____ TO _____		WORK UNIT NO.	
14. DATE OF REPORT (Yr., Mo., Day) 4/29/86		15. PAGE COUNT 15	
16. SUPPLEMENTARY NOTATION			
17. COSATI CODES			
18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)			
19. ABSTRACT (Continue on reverse if necessary and identify by block number)			
The report describes the results of a four year research effort to develop acoustooptic information processing systems capable of processing two dimensional signals. The general methodology that was developed for accomplishing this, is the use of a combination of temporal and spatial integrations in the optical system. Acoustooptic processors for synthetic aperture			
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS <input type="checkbox"/>		21. ABSTRACT SECURITY CLASSIFICATION	
22a. NAME OF RESPONSIBLE INDIVIDUAL		22b. TELEPHONE NUMBER (Include Area Code)	
		22c. OFFICE SYMBOL nm	

radar image formation were designed and experimentally demonstrated. The radar imaging processor is capable of forming images in real time with relatively low size and power requirements. Several processors were also developed for pattern recognition applications and an acoustooptic system capable of producing two dimensional correlations of images at standard video rates was experimentally demonstrated. The same signal processing method was applied to two dimensional spectrum analysis and the processing of signals from broadband adaptive phased arrays. In both instances acoustooptic architectures were developed and initial experimental demonstrations were performed.

Acousto-optic Processing
of 2-D Signals
Using Temporal
and spatial
integration

DISCLAIMER NOTICE

**THIS DOCUMENT IS BEST QUALITY
PRACTICABLE. THE COPY FURNISHED
TO DTIC CONTAINED A SIGNIFICANT
NUMBER OF PAGES WHICH DO NOT
REPRODUCE LEGIBLY.**

2

ACOUSTOOPTIC PROCESSING OF 2-D SIGNALS
USING TEMPORAL AND SPATIAL INTEGRATION

SUMMARY

The report describes the results of a four year research effort to develop acoustooptic information processing systems capable of processing two dimensional signals. The general methodology that was developed for accomplishing this, is the use of a combination of temporal and spatial integrations in the optical system. Acoustooptic processors for synthetic aperture radar image formation were designed and experimentally demonstrated. The radar imaging processor is capable of forming images in real time with relatively low size and power requirements. Several processors were also developed for pattern recognition applications and an acoustooptic system capable of producing two dimensional correlations of images at standard video rates was experimentally demonstrated. The same signal processing method was applied to two dimensional spectrum analysis and the processing of signals from broadband adaptive phased arrays. In both instances acoustooptic architectures were developed and initial experimental demonstrations were performed.



By _____	
Distribution/ _____	
Availability Codes	
Dist	Avail and/or Special
A-1	

Time-and-space integrating processing

A primary motivation for this work has been the availability of well developed acoustooptic devices and the desire to use these devices in performing two dimensional processing. Traditionally, optical image processing was done exclusively with two dimensional spatial light modulators and the lack of well developed devices of this type has long been considered a major impediment for progress in optical image processing. If we could find effective ways for processing images using well developed acoustooptic devices, then this can open the way for rapid progress to take place.

Acoustooptic devices are one dimensional spatial light modulators and in order to use them in an image processor it is necessary to find ways for representing and processing 2-D signals with one dimensional devices. The method we developed involves a combination of temporal and spatial representation of the signals. Accordingly, the processing of the data is accomplished by a combination of temporal and spatial integrations in the optical system, hence we refer to this method as time-and-space integrating processing. In all these processors, both transverse spatial dimensions of the optical system are used. There is great deal of flexibility in designing such architectures, resulting from the choices that exist on how to combine spatial and temporal integrations and how to arrange one dimensional devices in the three dimensional optical

system. We have explored these architectural possibilities in depth. The results of this investigation have been summarized in a review article listed as Publication #1 in this report.

The primary result from our investigation on two dimensional acoustooptic architectures is that it is possible to implement a wider class of 2-D signal processing operations using one dimensional, acoustooptic devices rather than two dimensional spatial light modulators. Moreover, it is often possible to electronically control the operation being performed in a very natural way, in the acoustooptic image processing architectures making a degree of programmability possible. On the negative side, since processing is partially done by temporal integration, the output signal forms on top of a bias, a characteristic of all time integrating processors. A consequence of this bias built-up is reduced dynamic range. We have investigated several methods for reducing the bias in the architectures we have considered and we have concluded that in almost all cases, if the system is properly designed, the dynamic range that is obtainable is sufficient for practical applications.

Four separate image processing application areas were selected for which specific architectures were developed and experimentally demonstrated: synthetic aperture radar, image correlation, two dimensional spectrum analysis, and adaptive, broadband phased arrays. The results from our investigation in

each of these areas have been previously published. In what follows we will highlight our results in each area and identify the relevant references where more detailed descriptions can be found.

Synthetic Aperture Radar

Synthetic Aperture Radar is the oldest and still the most successful application of optical signal processing techniques. Also it is a two dimensional signal processing problem and hence it was the most obvious problem to which our acoustooptic processing method could be applied. The advantage of using acoustooptics versus photographic film, is real time operation a relatively compact and power efficient system. The details of the operation of the acoustooptic SAR processor, several modified architectures, and the experimental demonstration, have been documented in publications #5, #6, and #7. Here we present a sample of the experimental results. The system block diagram is shown in Figure 1a. The processor accepts the radar signal and produces at its output the focused radar image as a video signal which can be readily displayed on a monitor. A photograph of the experimental set-up is shown in Fig.1b. The focused image of a simulated point target produced in real time by the acoustooptic system is shown in Figure 1c.

Several modifications to the basic SAR architecture were implemented allowing the imaging operation to be controlled electronically in order to accommodate dynamic changes in the

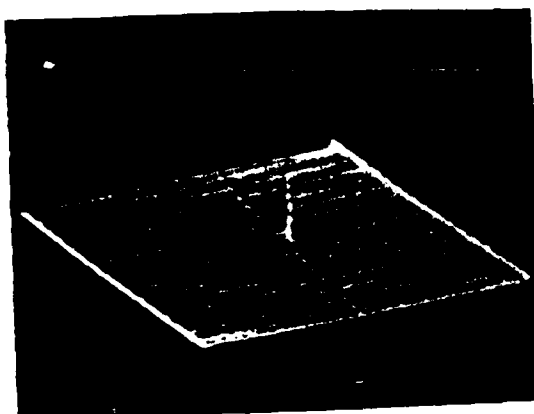
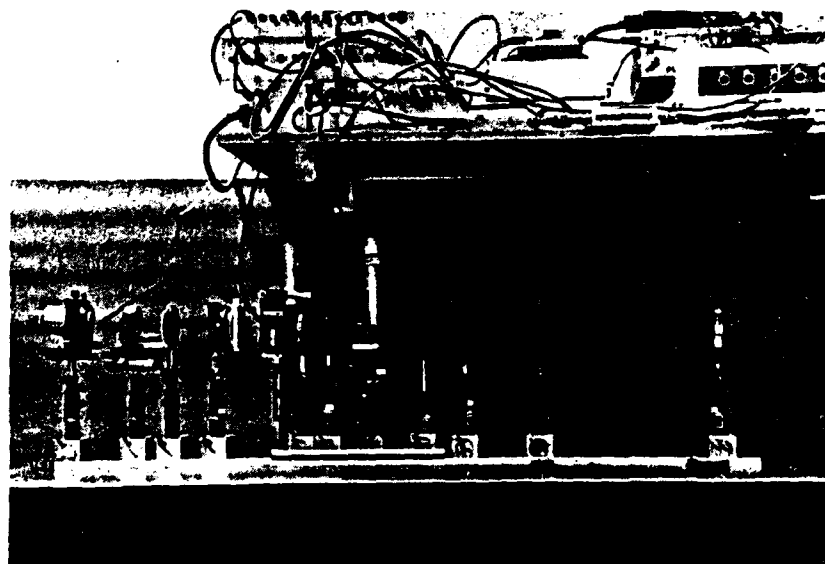
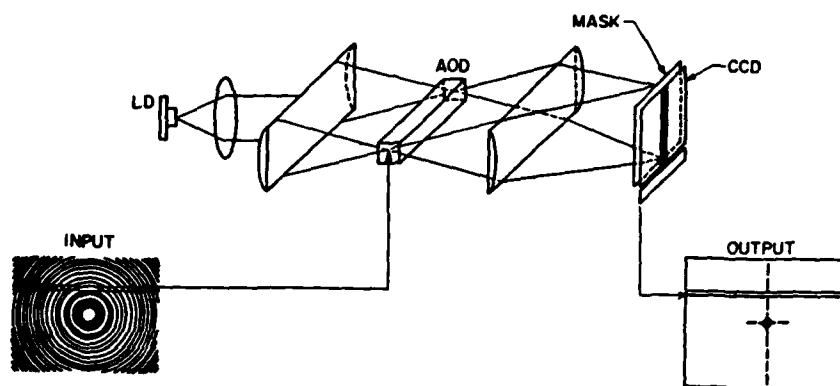


Figure 1. Acoustooptic Synthetic Aperture Radar Processor. a) Block Diagram b) Experimental Apparatus c) Focused Image of a simulated Point Target.

radar geometry. Most significantly, in publications #6 and #7 we report on an architecture in which the two dimensional mask that is used in the system of Fig. 1 is replaced by a second acoustooptic device.

Acoustooptic Image Correlators

Optical image correlation for pattern recognition applications was proposed by VanderLugt over 20 years ago. This technique has the potential for the most wide applicability of optical signal processing methods. There are two basic limitations that have so far prevented optical correlators from being used in practice: lack of two dimensional devices and the limitations of correlation based pattern recognition algorithms. The use of acoustooptic devices in place of two dimensional spatial light modulators obviously addresses the first problem. An acoustooptic image correlator was capable of processing images at 30 frames per second (or faster) was demonstrated. This system is described in detail in publication #2 and initial experimental results are presented in publication #3. In Figure 2 we present more recent experimental results from this processor. Shown in figure 2 are the auto and cross correlations between two selected images produced at video rates by the acoustooptic correlator.

The excellent quality of the results we obtained with the acoustooptic image correlator, have led us to the conclusion that the implementation limitation of optical correlators cannot

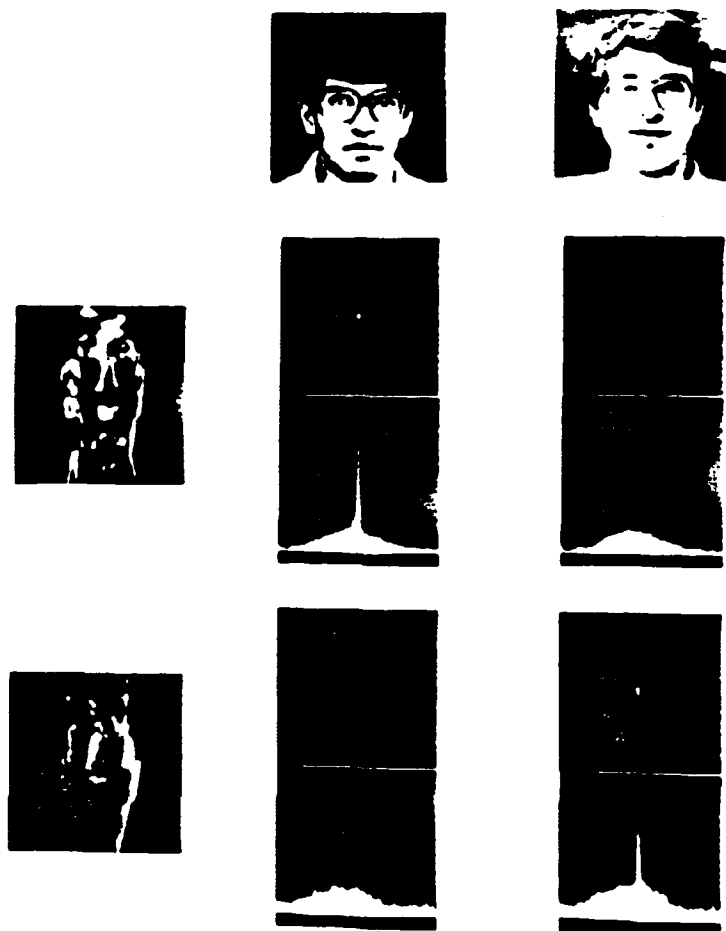


Figure 2. Correlations produced in real time by the experimental acoustooptic image processor.

be considered as a major problem preventing optical correlators from being used in practical pattern recognition applications. It is the limitations of the simple correlation algorithm that need to be overcome now. To this end we have investigated architectures where the reference image in the correlator can be digitally programmed. This capability opens new algorithmic possibilities that show promise for overcoming the limitations of the correlation algorithm. The architecture described in publication #4 is an acoustooptic image correlator in which the reference image is stored in an electronic memory and thus it can be directly updated by an electronic digital computer.

Spectrum Analysis

Two dimensional spectrum analysis has two basic applications. The first is the formation of the spectrum of inherently two dimensional signals (images). It is known however that the 2-D transform of a raster recorded 1-D signal, is approximately equivalent to the raster recording of the 1-D transform of the same signal. Hence, a 2-D spectrum analyzer can be used for producing very fine resolution spectra of one dimensional signals over a broad bandwidth. 2-D acoustooptic spectrum analyzers are actually better suited for this latter application because they can accept the one dimensional signal directly and the "raster recording" step occurs automatically during the computation of the spectrum. Several architectures were developed and compared. These are described in publication #8 along with initial experimental results.

Our investigation in this area is not complete but there are some general observations that we can make at this stage. The first is that the electronic circuits that are necessary for generating the signals that are required for the operation of the acoustooptic spectrum analyzer are relatively complex. Secondly, the dynamic range reduction due to the bias built-up can be a particularly serious problem for spectrum analysis. Thus our efforts are now concentrated towards solving these two problems.

Adaptive Phased Arrays

Processing signals from broadband, antenna arrays is a two dimensional signal processing problem; we have the time history of an array of signals, one from each antenna element. For the narrowband case, beamforming (i.e, spatial processing) and temporal filtering can be done independently as two cascaded operations. In the broadband case however temporal and spatial processing are inherently coupled and the problem must be treated as two dimensional. There are several distinctions between this two dimensional signal processing problem and the problems discussed earlier. These distributions translate to marked differences in the optical architectures that are needed for solving these problems. The input signal in this case an array of temporal signals which has led us to consider multiple transducer acoustooptic devices as input spatial light modulators. Another major difference is that for adaptive processing, the operation required is a cascade of two 2-D filters; the output of one filter becomes an input to the second. Finally there is feedback

which raises the issue of stability. These considerations have led to very interesting new architectural designs. These are described in publications #11 and #12. Also presented in these publications are initial experimental results from these architectures.

As described in the publications mentioned above, there are two basic architectural choices. Either a fully space integrating system, i.e. both required filters are performed using a space integrating architecture or a hybrid time/space integrating architecture. The advantage of the fully space integrating system is that the active devices required for its implementation, acoustooptic devices and detectors, are well developed and readily available. In addition, the space integrating architectures have fast adaptation time. The disadvantage of the space integrating architectures is susceptibility to oscillations. The problem with building a time/space integrating adaptive processor on the other hand, is that the output of the time integrating system needs to become the input to the space integrating processor (see publication #12). Normally, the output in a time integrating processor forms on a detector array. The requirement for electronically reading out the array before driving the next stage, would render this approach impractical. We have investigated the use of photorefractive crystals as a time integrating medium. The result of the temporal integration in this case is stored in the crystal and it can be optically addressed. Consequently, the time and space integrating systems

can be directly cascaded. This technique is discussed in publication #20. This method has yielded what we feel are excellent results, and it is applicable not only to this adaptive filtering problem, but it creates the possibility for a new class of acoustooptic architectures.

Graduate Students Associated with this program.

Kelvin Wagner: He worked on two dimensional acoustooptic architectures, synthetic aperture radar, and spectrum analyzers. September 1981-present
Master's degree in Electrical Engineering 1982.
Ph.D. degree in Electrical Engineering Aug. 1986 (expected).

Michael Haney: He worked on synthetic aperture radar. 1981-1986.
Ph.D. degree in Electrical Engineering February 1986.
Present employment: General Dynamics, Pomona.

Hyuk Lee: He worked on Acoustooptic devices, 1981-1985.
Ph.D. in Applied Physics, December 1985.
Present employment: Assistant Professor at Polytechnic Institute of New York.

Santosh Venkatesh: He worked on image processing algorithms and their implementation with acoustooptic devices, 1981-present.
Master's degree in E.E. 1982.
Ph.D. degree in E.E. July 1986, expected.

List of Publications that resulted from the program

- (1) "Two-Dimensional Optical Processing Using One-Dimensional Input Devices", Demetri Psaltis, Proceedings of the IEEE, Vol. 72, No. 7, July 1984.
- (2) "Optical image correlation using acoustooptic and charge-coupled devices", Demetri Psaltis, Applied Optics, Vol. 21, No. 3, 1 February 1982.
- (3) "Acousto-Optic/CCD Image Processor", Demetri Psaltis, Eung Gi Paek and Santosh Venkatesh, Proceedings of the IOCC Conference, Boston, 1983.
- (4) "Incoherent Electro-optic Image Correlator", Demetri Psaltis, Optical Engineering, Vol. 23 No. 1, January/February 1984.
- (5) "Real-time Optical Synthetic Aperture Radar (SAR) Processor", Demetri Psaltis and Kelvin Wagner, Optical Engineering, Vol. 21, No. 5, September/October, 1982.
- (6) "Programmable Real-Time Acousto-Optic/CCD SAR Processor", Michael Haney, Kelvin Wagner, Demetri Psaltis, SPIE Proceedings #495-25.
- (7) "Acousto-optic Techniques for Real Time SAR Imaging", Michael Haney, Demetri Psaltis, SPIE Proceedings, #545-26,
- (8) "Time and Space Integrating Acousto-optic Folded Spectrum Processing for SETI", Kelvin Wagner and Demetri Psaltis, SPIE Proceedings #564-31,
- (9) "Real Time Computation of Moments with Acousto-optics", Kelvin Wagner and Demetri Psaltis, SPIE Proceedings #352-19,
- (10) "A Space Integrating Acousto-Optic Matrix-Matrix Multiplier" Kelvin Wagner and Demetri Psaltis, Optics Communications Vol. 52 No. 3, 1 December 1984.
- (11) "Adaptive Acoustooptic Processor" Demetri Psaltis and John Hong, SPIE Proceedings #519-09,
- (12) "Acousto-optic Adaptive Signal Processing", John Hong and Demetri Psaltis, SPIE Proceedings #551-20,

- (13) "Measurement of the temporal coherence properties of pulsed single-mode laser diodes", Michael Haney and Demetri Psaltis, Applied Optics, Vol 24, No. 13, 1 July 1985.
- (14) "Acousto-Electro-Optic Light Modulation", Demetri Psaltis, Hyuk Lee and Gabriel Sirat, SPIE Proceedings #465-23,
- (15) "Acousto-electro-optic light modulation" Demetri Psaltis, Hyuk Lee, and Gabriel Sirat, Applied Physics Letters, Vol. 46, No. 3, 1 February 1985
- (16) "Photorefractive incoherent-to-coherent optical converter", Y. Shi, D. Psaltis, A. Marrakchi, and R. Tanguay, Jr., Applied Optics, Vol. 22, No. 23, 1 December 1983.
- (17) "Physical characterization of the photorefractive incoherent-to-coherent optical converter", A. Marrakchi, A. R. Tanguay, Jr., J. Yu, and D. Psaltis, Optical Engineering, Vo. 24, No. 1, January/February 1985.
- (18) "Photorefractive Incoherent-To-Coherent Optical Conversion", D. Psaltis and J. Yu, SPIE Proceedings #465-01, 1984.
- (19) "Photorefractive Incoherent-To-Coherent Optical Converter: Physical and Materials Considerations", A. Marrakchi, A. R. Tanguay, Jr., J. Yu, and D. Psaltis, SPIE Proceedings #465-12,
- (20) "Bias-free time-integrating optical correlator using a photorefractive crystal", Demetri Psaltis, Jeffrey Yu, and John Hong, Applied Optics, Vol. 24, No. 22, 15 November 1985.
- (21) "Optical implementation of the Hopfield model", Nabil H. Farhat, Demetri Psaltis, Aluizio Prata, and Eung Paek, Applied Optics, Vol. 24, No. 10, 15 May 1985.

Two-Dimensional Optical Processing Using One-Dimensional Input Devices

DEMETRI PSALTIS, MEMBER, IEEE

Invited Paper

Two-dimensional optical processing architectures that are implemented with one-dimensional input spatial light modulators are reviewed. The advanced state of the art of available one-dimensional devices and the flexibility that exists in the design of two-dimensional architectures with one-dimensional transducers leads to the implementation of the most powerful and versatile optical processors. Signal and image processing architectures of this type are discussed.

I. INTRODUCTION

Optical information processing systems can be configured in two or three spatial dimensions, one of the dimensions being the optical axis of the system along the general direction in which light propagates. As light propagates through the system it is modulated by input data and transformed to produce, at the output plane, a light distribution modulated by the processed data (the result of the computation). Light modulators and optical processing elements are typically placed in planes perpendicular to the optical axis. An optical processor is referred to as one- or two-dimensional (1-D or 2-D) depending on whether one or two of the transverse coordinates of the optical system are used. The majority of the early optical information processing systems were 2-D processors [1]–[3]. The 2-D space provides large parallel processing capability that gives optical processors the potential for extremely high processing power. In addition, the inherent 2-D processing capability makes optics a signal processing technology that is uniquely suited for image processing applications, which typically require extensive computations. Indeed, the first successful application of optical information processing systems was in the area of image processing, particularly in synthetic aperture radar [2] and pattern recognition [3]. 1-D processors were developed subsequently, primarily due to the emergence of acoustooptic devices (AODs) as broadband, high-quality electronic-to-optical transducers [4]. The acoustooptic spectrum analyzer is the most widely used processor of this type, consisting simply of an AOD, a 1-D Fourier transforming lens, and a linear detector array [5]. In recent years, a new generation of 2-D optical architectures has emerged. These architectures are implemented with

1-D input devices but are configured in three spatial dimensions. The classical 2-D processors, which are implemented with 2-D spatial light modulators (SLMs) [6], [7], process data in a relatively straightforward manner, typically using spherical lenses to perform Fourier transformations or 2-D shift-invariant operations. The implementation of 2-D processors with 1-D devices is usually less straightforward, but greater flexibility exists in designing such systems. Consequently, many new architectures have emerged in a relatively short time, in which broad-band, high-quality 1-D devices have been incorporated in the 2-D optical system to produce the most powerful and versatile optical computers to date. We will examine in this paper the methods that have been proposed for implementing such processors. Before we proceed with descriptions of specific architectures we will discuss in this introductory section general characteristics of processors of this type.

The most direct motivation for using 1-D devices is their advanced state of the art. In recent years, numerous optical devices have been developed, usually for applications other than optical information processing. Nevertheless, many of these devices have been effectively used as input or output transducers in 1-D optical processors. AODs, semiconductor light sources (laser diodes and LEDs) [8], and semiconductor detectors (photodiodes and CCDs [9], [10]) are the components that have been most widely used in this manner. The relatively small size of these devices combined with advanced computer-aided lens design techniques has led to the development of miniaturized optical systems [11]. In addition, the relatively low power consumption and cost of the devices has made it possible to implement optical processors that are advantageous not only in terms of speed but also in terms of power requirement, size, and cost. These practical considerations provide strong impetus for using these same devices to implement 2-D processors with similar properties.

The ability to configure the processor in three dimensions is a property unique to optics among signal processing technologies. It is crucial to utilize the third dimension effectively because, in general, it is this property that gives optics a clear advantage over alternate technologies. Therefore, an important question is whether 2-D optical processors that are configured with 1-D devices can have sufficient processing power, because the full potential of the three-dimensional optical system can only be realized with the classical implementation of a 2-D processor utilizing 2-D SLMs. If such a processor is configured to perform

Manuscript received January 5, 1984. Research on optical architectures at Caltech is supported by the Air Force Office of Scientific Research, the Army Research Office, NASA, and General Dynamics. The author is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125, USA.

2-D correlations, then each of 10^6 samples that are stored on a 2-D SLM at the input plane, is multiplied by 10^6 samples stored on a second 2-D SLM placed in the spatial frequency plane. This massive parallelism and interconnection capability results in a processing rate equivalent to 10^{14} analog multiplications per second, assuming that the input SLM is updated at 100 frames per second. When the input device is 1-D, some of this parallelism is lost since only 10^3 – 10^4 samples can now be represented in the input plane, at one time. The loss in parallelism, however, is compensated for by the very high bandwidth of the available 1-D devices. For instance, consider a processor that is implemented with a linear array of 100 laser diodes as the input device. Each laser diode can be separately modulated with a bandwidth equal to 1 GHz, and therefore the input data can be updated every 1 ns. The processing power in this case can exceed 10^{14} analog multiplications per second, if the light from each laser diode is multiplied by at least 10^3 samples in the optical system. It is actually possible to multiply each laser diode with up to 10^6 samples and thus, given the properties of the devices that are available, larger processing power can be obtained with a 2-D optical processor if it is implemented with 1-D devices. Notice that in the above example, the 10^3 interconnections are only a small fraction of the interconnection capacity of the 2-D optical system, yet they are sufficient to yield very high processing power due to the use of the broad-band input devices.

A consequence of the reduced utilization of the interconnection capacity of the optical system is increased flexibility. The 2-D Fourier transforming lens is the primary mechanism that is used to make all the possible interconnections in a 2-D optical processor. However, only a relatively small number of linear operations can be performed with a processor based on the 2-D Fourier transform. It is in fact possible to compute a wider class of linear operations through combinations of imaging and 1-D transforming (or integrating) lenses. Furthermore, greater flexibility exists in the design of such processors, permitting the optimization of the optical architecture so that a particular operation is performed in a manner that is matched to the requirements of each application.

A second important consequence of the use of broad-band transducers has been the development of "dynamic" 2-D optical processors. Time plays a passive role in a classical 2-D processor: data are placed at the input plane, they are processed by spatial integration in the optical system, and then a time interval elapses during which the last result is read out and a new block of data is entered in the optical processor. The incorporation of devices such as AODs and CCDs, which have the capability of transferring data not only into, but also through the optical system, results in *dynamic* optical processors in which different calculations are continuously being performed. Time integration is an example of dynamic optical processing [12], [13]. As data flow through a time integrating system, computations are continuously performed and the partial results are stored (accumulated) on the optical detector by buildup of photo-generated charge. Recently, a more deliberate utilization of the time domain has been achieved through systolic 1-D and 2-D optical processing [14]–[17]. Systolic optical processing is addressed in a separate paper in this issue. A byproduct of dynamic processing is increased programma-

bility in such systems. Usually, the operation performed is controlled by electrical signals externally applied and it is a relatively easy task to reprogram the processor by changing these signals with an electronic processor that is interfaced to the optical system.

In summary, practical as well as powerful 2-D optical processors can be implemented using 1-D input devices, primarily because of the advanced state of the art of the available components. Furthermore, such systems are particularly versatile due to the latitude that exists in designing the spatial configuration as well as the timing of these architectures. In the remainder of the paper we explore these issues further by discussing several processors of this type. The architectures that will be described are implemented with astigmatic optical systems, typically consisting of a combination of spherical and cylindrical lenses in conjunction with the active devices used in each case. Even within this relatively narrow framework, it is not possible to cover all the systems that have been proposed. Instead, several representative architectures were selected, partially because of their significance but also because they were judged to be suitable for conveying crucial concepts.

The processors that will be discussed are all implemented with AODs and arrays of laser diodes (LDs) or LEDs as input devices. There are other 1-D devices that can be used instead, such as the array of electrooptic modulators that was recently fabricated at Xerox [59] or the CCD spatial light modulator that is being developed at Lincoln Laboratory [60]. AODs, LEDs, and LDs are used throughout the paper because these devices have been primarily used in practice and also because they are representative of the two basic types of 1-D input devices. In a LED or LD array, each element of the array can be individually modulated, and thus an entire line of data can be entered into the optical processor, in parallel, in a time approximately equal to the inverse of the bandwidth of the individual devices (1 ns for LDs). On the other hand, data are entered serially into the optical system when an AOD is used. If the voltage applied to an AOD is temporally modulated by $f(t)$, then the light diffracted by the device is modulated by $f(t - x/v)$, where x is the spatial coordinate along the direction of propagation of the acoustic wave and v is the acoustic velocity. At any time, t , the diffracted light is spatially modulated proportionally to a portion of the electrical signal applied to the device; equivalently, at any position, x , the light is temporally modulated by the input signal. This dual modulation is important for the synthesis of dynamic optical processors because AODs can be used not only as input devices but also as delay lines in the optical processor.

The modulation introduced by the AOD is actually more complicated [4]: the diffracted light is Doppler shifted to the center frequency of the AOD, the finite aperture of the device places a limit on the spatial extent over which the modulation takes place, and the approximation that the modulation is proportional to the applied electrical signal is valid only if the diffraction efficiency of the device does not exceed a few percent [18]. Furthermore, the modulation properties of AODs depend strongly on the angle of incidence of the illuminating light, an aspect that becomes particularly important in the design of astigmatic systems since in these architectures AODs are often illuminated with uncollimated light [19]. All these factors must be taken into consideration in the design of a practical system. How-

ever, since our purpose in this paper is to describe the functional operation of 2-D processor architectures, the simple form of the modulating function, $f(t - x/v)$, will be used except in cases where the additional factors play a vital part in the architecture. In architectures that are implemented with incoherent light or where operations are performed on detected signals (time integrating processors), bipolar signals are typically placed on a bias so that they can be represented by light intensity, which is a positive quantity. For clarity, the details of biasing in such systems will be suppressed.

The rest of the paper is divided in three sections. The following section examines multichannel systems. 2-D optical processors for processing 1-D signals are discussed in Section III, and the final section is devoted to optical processors capable of processing 2-D signals (images).

II. MULTICHANNEL PROCESSORS

The basic linear operations (Fourier transform, convolution, and correlation) can be implemented with 1-D optical processors in a variety of ways [5], [20]. The majority of these architectures are (or at least can be) implemented using AODs as electronic-to-optical transducers. Such systems are successfully used in practice even though they are in direct competition with other analog and digital 1-D signal processing technologies, primarily because of the high bandwidth of AODs, which can exceed 1 GHz. The result is a processing rate that cannot be easily achieved with alternate technologies. The processing power of a 1-D optical processor can be increased further by a factor of 10^2 to 10^3 by simply performing the same basic 1-D operation in all the available channels along the second dimension of the optical system. Such multichannel processors were proposed early, for instance by Cutrona in his classic paper in 1964 [1]. These systems are useful in applications where one wishes to process a signal in many different ways, or alternatively when the same processing operation is to be performed on a number of different signals simultaneously.

The system shown in Fig. 1 is an example of a multichannel 1-D space integrating correlator. The input signal $f(t)$ is applied to the AOD in Fig. 1. The AOD is illuminated by a coherent light source. The spherical lens L_1 is a collimating lens and the cylindrical lens L_2 focuses the light in the

vertical direction onto the AOD. The amplitude of the diffracted light is modulated by $f(t - x/v)$, where x is the horizontal spatial dimension and v is the acoustic velocity. The undiffracted light transmitted through the AOD is blocked in the focal plane of the spherical lens L_3 , whereas the diffracted light is collimated in the vertical dimension by L_3 . In the horizontal dimension, L_3 and the cylindrical lens L_4 produce an image of the diffracted light at plane P_2 . The light illuminating plane P_2 is therefore modulated by $f(t + x/v)$ and is uniform in the vertical (y) dimension. A 2-D transparency with amplitude transmittance $h(x, y)$ is placed at P_2 . The amplitude of the light transmitted through P_2 is modulated by the product $f(t - x/v)h(x, y)$. The combination of the cylindrical lens L_5 and the spherical lens L_6 image plane P_2 onto plane P_3 in the vertical dimension. In the horizontal dimension, L_6 produces the Fourier transform of the light exiting plane P_2 . The amplitude of the light at the output plane P_3 is modulated by

$$g(\omega_x, y, t) = \int f(t + x/v) h(x, y) \exp(j\omega_x x) dx \quad (1)$$

where $\omega_x = 2\pi x'/\lambda F_6$ is the radial spatial frequency, λ is the wavelength of light, F_6 is the focal length of L_6 , and x' is the horizontal coordinate at plane P_3 . If a slit is placed at P_3 , oriented vertically and centered on the optical axis, then light is transmitted through the slit only around $\omega_x = 0$. The width of the slit is chosen to be less than $\lambda F_6/A$ (A being the aperture of the mask at plane P_2) and a detector array is placed immediately after the slit. The photocurrent that is generated at a detector element located at position y , is proportional to

$$g(t, y) = \left| \int f(t + x/v) h(x, y) dx \right|^2 \quad (2)$$

The temporal modulation of the output signal from each detector is the magnitude squared of the correlation of the input signal $f(t)$ and the signal that is recorded on the mask as a function of position x at the position along the y dimension where the corresponding detector element is placed. The number of parallel correlations that this system can compute is limited by the size of the parallel readout detector array that can be fabricated—in practice approximately 100.

The multichannel, time integrating spectrum analyzer [21] shown in Fig. 2 is an example of a multichannel system that

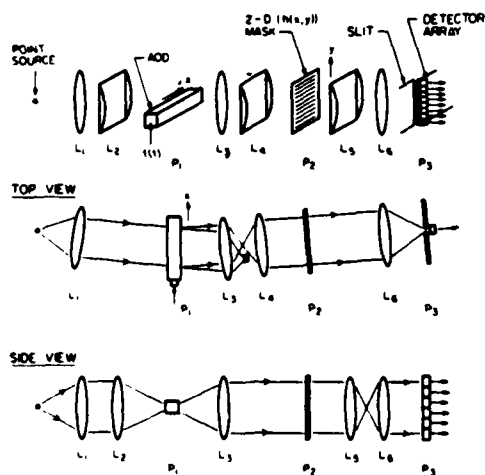


Fig. 1. Multichannel space integrating correlator.

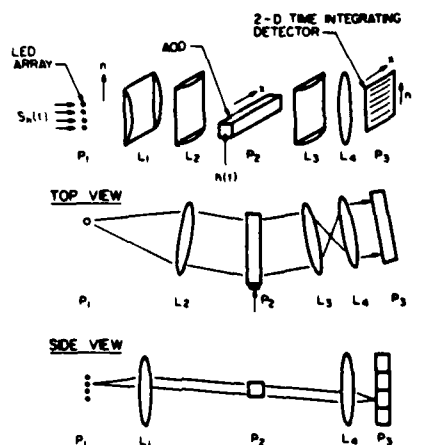


Fig. 2. Multichannel time integrating spectrum analyzer.

can accept many signals in parallel and perform the same operation on each. A linear array of LEDs is placed in the input plane P_1 of Fig. 2. The pair of cylindrical lenses L_1 and L_2 collimate the light from each LED so that they all illuminate the AOD placed in plane P_2 . The focal length of lens L_1 is chosen to be much shorter than the focal length of L_2 so that the beams illuminating the AOD are narrow in the vertical direction and wide horizontally matching the shape of the aperture of the AOD. The intensity of the n th LED is modulated by

$$I_n = I_0 + s_n(t) \cos(\omega_0 t + bt^2) \quad (3)$$

where $s_n(t)$ denotes the n th input signal, ω_0 is a fixed frequency the significance of which will be discussed shortly, b is a constant that has units s^{-2} , and I_0 is a bias that is added to each signal to allow negative values to be represented by light intensity. The modulating signal applied to the AOD is $h(t) = \cos(\omega_0 t + bt^2)$. When a bias is added to this signal before it is applied to the AOD, the intensity of the diffracted light is modulated proportionally to $h(t - x/v)$. Therefore, the intensity of the light diffracted by the AOD in Fig. 2 is modulated by the product of the incident intensity and $h(t - x/v)$. The cylindrical lens L_3 and the spherical lens L_4 form an image of this diffracted light in the horizontal (x) direction at the output plane P_3 . The lens L_4 focuses the light vertically onto the output plane or, equivalently, the combination of lenses L_1 and L_4 produce an image of the LED array in the vertical direction at the output plane. Consequently, light that originates at the n th LED is detected on the n th row of a 2-D detector array at plane P_3 . The photogenerated charge on a detector element located at position x at the n th row of the array is proportional to the time integrated intensity of the incident light:

$$\begin{aligned} I_D(x, n) &= \int I_n(t, x) h(t - x/v) dt + \text{bias terms} \\ &= \int s_n(t) \cos(\omega_0 t + bt^2) \\ &\quad \cdot \cos(\omega_0 t + b(t - x/v)^2) dt + \text{bias} \\ &= (1/2) \int s_n(t) \\ &\quad \cdot \cos(\omega_0 x/v + 2bxt/v + bx^2/v^2) dt + \text{bias} \\ &= (1/2) |S_n(2bx/v)| \\ &\quad \cdot \cos[\omega_0 x/v + \phi_n(2bx/v) + bx^2/v^2] + \text{bias}. \end{aligned} \quad (4)$$

In the above equation

$$S_n(\omega) = |S_n(\omega)| \exp[\phi_n(\omega)] = \int s_n(t) e^{-j\omega t} dt$$

is the Fourier transform of the real signal $s_n(t)$. Each transform is recorded as a function of position x at a different row of the detector array and therefore the spectra of all input signals can be independently accessed. The magnitude of each transform modulates the amplitude of the spatial carrier in (4) whereas the phase of the transform appears as phase modulation on the same carrier. Both quantities can be obtained directly when the detector is read out, by electronically filtering the detector signal to remove the bias terms in (4) and mixing the filtered signal with $\cos(\omega_0 x + bx^2/v^2)$ in quadrature to obtain the real

and imaginary components of $S_n(\omega)$. This Fourier transforming algorithm, known as the "chirp-z" in the digital signal processing literature [22], has proven very useful in optics since it allows the calculation of the transform of long signals with fine resolution, using temporal integration on the detector. The integration time is limited by the dark current of the detector to several milliseconds or longer if the detector is cooled. In general, the system of Fig. 2 can be used for the calculation of the spectra of many relatively narrow-band signals with fine resolution. The number of parallel channels in this case is limited by the number of LEDs (or laser diodes) that is feasible to assemble and drive in parallel. LD arrays have been fabricated monolithically with 18 elements at 15- μm centers [23] and it is feasible to fabricate larger arrays either monolithically or by stacking individual modules.

III. 2-D PROCESSING OF 1-D SIGNALS

In this section we consider optical processors designed to process 1-D signals in both transverse spatial dimensions. Generally, more complex 1-D signal processing operations can be performed by using the two dimensions directly, resulting in optical processors that are not only computationally more powerful than 1-D processors (as are multichannel systems), but also more versatile and capable of performing a wider class of signal processing operations. In some cases the distinction between multichannel and 2-D processors is subtle. For instance, the multichannel system of Fig. 1 can also be thought of as processing its input signal in both spatial dimensions to perform a single linear operation. We will see, however, through the examples in this section that more direct utilization of the 2-D space is possible.

The first architecture considered is the vector-matrix multiplier shown in Fig. 3. This system demonstrates the

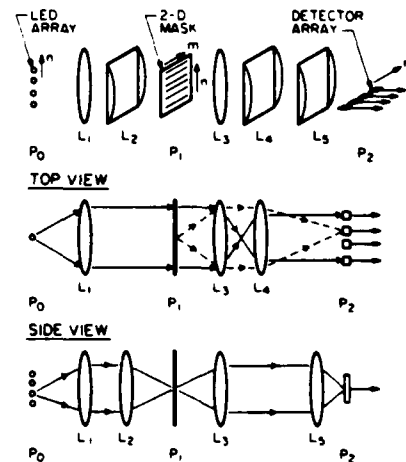


Fig. 3. Vector-matrix multiplier, implemented with an array of LEDs

impact that device developments have had on architectural design. The architecture in Fig. 3 was originally proposed in 1964 [1], but the full potential of this system was appreciated more than 10 years later [24]–[27], by which time broad-band source and detector arrays had become feasible. We examine here the implementation proposed by

Goodman *et al.* [24]. An array of LEDs is positioned vertically at the input plane P_0 in Fig. 3. The intensity of the light emitted by the n th LED is denoted by $A(n)$. The spherical lens L_1 collimates the light from each LED in the horizontal direction and the combination of L_1 and the cylindrical lens L_2 produce an image of the LED array at plane P_2 in the vertical direction. Consequently, the 2-D mask placed at P_2 is illuminated uniformly in the horizontal direction and each row of the mask is illuminated by a different LED. The intensity transmittance of the mask is denoted by $B(n, m)$, which signifies that a pixel of the mask with area $(Dx \cdot Dy)$, located at coordinates $x = m \cdot Dx$ and $y = n \cdot Dy$, has average transmittance $B(n, m)$. The mask is imaged in both dimensions onto the output plane P_3 . In the horizontal dimension, the spherical lens L_3 and the cylindrical lens L_4 produce a relatively large image of the mask so that each column of the mask is imaged onto a separate element of a linear detector array placed at P_3 . In the vertical dimension, L_3 and L_5 produce a demagnified image of the mask that is smaller in the vertical direction than the height of each detector element. Therefore, the photogenerated signal at the m th detector element is proportional to the total intensity transmitted through the m th column of the mask:

$$C(m) = \sum_n^N B(n, m) A(n) \quad (5)$$

where N is the number of LEDs. The operation performed in this case is the multiplication of the vector A , applied to the LED array, and the matrix B , stored on the mask. Equivalently, it is a 1-D linear operation with a space-variant kernel. The use of the second dimension of the optical system provides in this architecture the capability to perform a wider class of operations, since it is only possible to implement shift-invariant operations and the Fourier transform with 1-D systems. It is interesting to calculate the processing power of this relatively simple optical processor. The size of the vector and the matrix that can be processed is limited by the number of elements in the LED and detector arrays. For arrays with 100 elements, 10^4 analog multiplications are performed by the optical system each time a new vector is entered and the product read out by the detector. If the LEDs and the detectors have a 10-ns response time, the resulting processing rate is 10^{12} analog multiplications per second. The fact that this large processing power can be obtained with such a simple optical system, with which computations can be performed with relatively high accuracy, has raised the possibility of performing numerical operations optically.

Vector-matrix products can also be formed with the architecture shown in Fig. 4 that has been proposed by

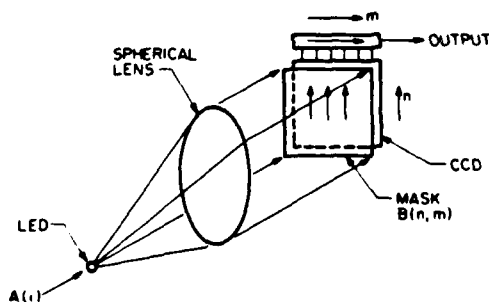


Fig. 4. Vector-matrix multiplier, implemented with a single LED and a 2-D CCD detector

Monahan *et al.* [28]. In this case a single LED is used that is collimated by a spherical lens to uniformly illuminate a 2-D transparency with intensity transmittance $B(n, m)$, where n and m are integers denoting the position of each pixel of the mask in the two spatial dimensions. The intensity of the LED is successively modulated by the elements of the vector $A(i)$, where i is an integer. The intensity of the light transmitted through the mask is modulated by the product $A(i)B(n, m)$. The light is detected by a 2-D CCD array placed immediately after the mask. There is a one-to-one correspondence between the pixels of the mask and the detector elements of the CCD and therefore the charge that is photogenerated at the (n, m) detector element when the i th element of the input vector is applied to the LED is $C(n, m, i) = A(i)B(n, m)$. The charge pattern stored in the CCD is transferred vertically by one pixel before the next element of the vector is applied to the LED. The charge that is photogenerated during the time the LED is modulated by the i th element of the input vector is transferred on the CCD by $N - i$ pixels, in the vertical direction, after all N elements of the input vector have been entered in the processor. The total charge that is accumulated on the CCD after the N elements of the vector are applied to the LED, is

$$C(n, m) = \sum_i^N B(n - N + i, m) A(i). \quad (6)$$

The CCD has n rows in the vertical direction and the top (N th) row is a CCD register that transfers its contents horizontally to the output port of the device. The charge accumulated at the top row is therefore the output signal of this processor and it is found by substituting $n = N$ in (6)

$$C(N, m) = \sum_i^N B(i, m) A(i). \quad (7)$$

Thus this system also produces at its output the product of the vector that modulates sequentially the intensity of the LED and the matrix stored in the mask. This example demonstrates how the same basic operation can be performed in different ways in a 2-D processor, each having distinct operational characteristics. The single LED and the 2-D detector array used in this architecture allow serial loading of data in the optical system, a convenient property in many applications. The parallel addressing of the system in Fig. 3, on the other hand, is useful in applications where the data naturally appear in parallel format, such as in processing signals from antenna arrays. Furthermore, the parallel input/output format of the system in Fig. 3 permits the implementation of iterative algorithms by feeding back the output signal from each detector element to the corresponding LED at the input plane of the processor [29].

Both architectures discussed thus far in this section require a 2-D transparency in addition to the input devices that are used to enter broad-band signals into the optical processor. This 2-D transparency is typically photographic film on which the matrix is permanently stored. Consequently, the matrix cannot be dynamically changed. A 2-D transparency is not, however, the only method that provides access to the 2-D space of the optical system. 2-D optical processors with dynamic spatial impulse response can be implemented using two 1-D modulators oriented perpendicular to one another. The first architecture of this type was reported by Said and Cooper [30]. The processor shown schematically in Fig. 5 is a modified version of the system that was originally proposed, in that two separate

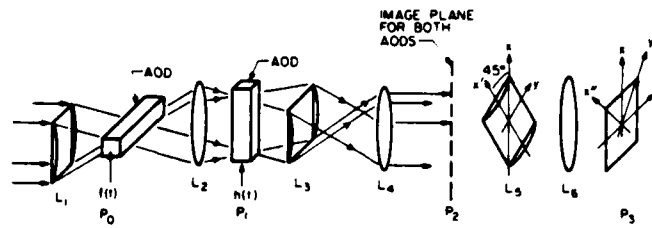


Fig. 5. Space integrating ambiguity function processor.

AODs are used here in place of a single device with two acoustic waves propagating in orthogonal directions. The monochromatic plane wave illuminating this system is focused in the vertical dimension by the cylindrical lens L_1 onto an AOD placed in the horizontal (x) direction. A signal $f(t)$ is applied to the AOD and the light diffracted by the device is recollimated vertically and focused horizontally by the spherical lens L_2 onto plane P_1 where a second AOD is placed oriented in the vertical (y) dimension. This AOD is driven by the signal $h(t)$. In the horizontal dimension, the doubly diffracted light is collimated by the spherical lens L_4 and the combination of lenses L_2 and L_4 form an image of the light diffracted from the first AOD at plane P_2 . In the vertical direction, the cylindrical lenses L_3 and L_4 form the image of the light diffracted by the second AOD at P_2 . The amplitude of the 2-D light distribution at P_2 is therefore modulated by $f(t - x/v)h(t - y/v)$. The astigmatic pair of lenses L_5 and L_6 image P_2 along one dimension and form the Fourier transform in the orthogonal dimension. The direction in which the cylindrical lens L_5 has focusing power is at 45° with respect to the x and y axes and therefore the Fourier transformation is performed along a direction that is also at 45° with respect to the (x, y) axes. It is convenient at this point to define a new set of coordinates, y' , parallel to the axis of the cylinder L_5 and x' perpendicular to y' . The old coordinates are related to the new by

$$\begin{aligned} x' &= (x + y)/\sqrt{2} \\ y' &= (x - y)/\sqrt{2} \end{aligned}$$

We can now readily write an expression for the amplitude of the light distribution at the output plane P_3 as the 1-D Fourier transform of the light distribution at plane P_2

$$g(u, y', t) = \int f[t - (x' + y')/\sqrt{2}v] h[t - (x' - y')/\sqrt{2}v] \cdot \exp(2\pi i x'u) dx' \quad (8)$$

where $u = x''/\lambda F_6$, x'' is the spatial coordinate at the output plane and F_6 is the focal length of L_6 . The function g in the above equation has some interesting properties. If $f(t) = h(t - \tau)\exp(j2\pi u_0 t)$ is a delayed version of the signal $h(t)$, shifted in frequency by u_0 hertz, then at $x'' = u_0 \lambda F_6 / \sqrt{2} v$ and at a time t when both f and h are within the apertures of the AODs, g as a function of position y' is the autocorrelation of h . The peak of this autocorrelation function occurs at $y' = v\tau/\sqrt{2}$. This operation, known as the ambiguity function [31], is useful in radar signal processing since the range (which is proportional to τ) and the velocity of a target (proportional to u_0) can be estimated from the location of the peak of g in the (u, y') plane. If the signals f and h are longer than the acoustic delay through the AODs, the ambiguity function of a different portion of the two signals is calculated at any one time. These partial

ambiguity surfaces can be incoherently accumulated in time on a 2-D detector at the output plane to enhance the signal-to-noise ratio. This architecture exemplifies the distinction between multichannel and 2-D processors. The operation performed along either of the two spatial coordinates is not independently part of the final answer, only the combined processing along both transverse dimensions produces the full ambiguity surface.

The system described above is interesting in its own right, but perhaps more importantly it demonstrated for the first time a very flexible and powerful method for processing data in two dimensions: the use of two orthogonal AODs or other 1-D devices. In recent years, many configurations have been proposed using this technique [32]–[36] and it is likely that more architectures of this type will be invented. We will examine here, as an additional example, a particularly flexible architecture proposed by Kellman [37] and Turpin [38] and shown in Fig. 6. A LED is the light

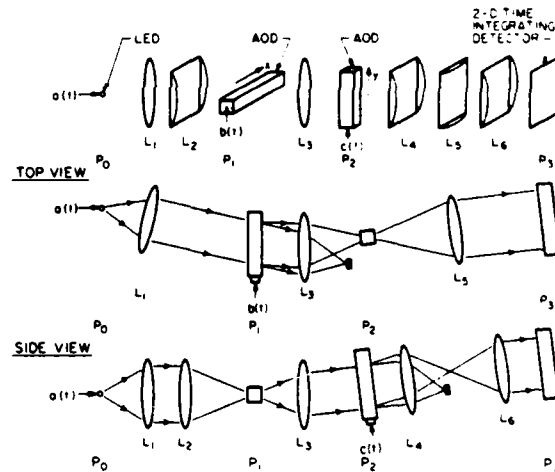


Fig. 6. Time integrating triple-product processor

source in this system, its intensity temporally modulated by the signal $a(t)$. The first AOD, oriented horizontally in Fig. 6, is modulated by the signal $b(t)$ and illuminated by the LED through lenses L_1 and L_2 . The intensity of the light diffracted by the AOD, modulated by the product $a(t)b(t - x/v)$, is expanded vertically and focused horizontally by L_3 to illuminate the second, vertically oriented AOD in Fig. 6. The signal applied to the second AOD is denoted by $c(t)$. The doubly diffracted light is modulated by the product $a(t)b(t - x/v)c(t - y/v)$. The cylindrical lenses L_4 and L_6 produce an image of the second AOD in the vertical direction at the output plane, while L_3 and L_5 image the first AOD in the horizontal direction onto the output plane. A 2-D detector array is placed at the output plane. The photogenerated charge that accumulates at a detector ele-

ment located at coordinates (x, y) and exposed to light for T seconds is proportional to

$$d(x, y) = \int_T a(t) b(t - x/v) c(t - y/v) dt. \quad (9)$$

This system is known as the triple product processor because the 2-D output function, $d(x, y)$, is obtained by temporally integrating the product of the three input signals. We can see from the top and side views in Fig. 6 that this processor consists of two 1-D time integrating correlators, operating simultaneously in orthogonal directions. Note the difference between this system and the multi-channel time integrating processor we discussed in the previous section (Fig. 2). A correlation is performed in this case in both dimensions; the result is obtained from the combined 2-D processing rather than the repetition of the 1-D processing operation. The operation performed by this architecture always has the general form given by (9), but there is considerable flexibility since all three input signals are electronic: by altering them new operations can be performed with the same optical hardware. For instance, if $a(t) = s(t) \exp(jkt^2)$, $b(t) = s(t)$, and $c(t) = \exp(-jkt^2)$, where k is a constant, then the output, $d(x, y)$ becomes

$$d(x, y) = \exp[-jk(y/v)^2] \int s(t) s(t - x/v) \cdot \exp[-2j(ky/v)t] dt. \quad (10)$$

The above expression is the ambiguity function of the signal $s(t)$ multiplied by a quadratic phase term (which can be removed, if necessary, by post-detection multiplication). Notice that in (10) we have allowed the input signals to be complex. In practice, complex signal processing capability is achieved by modulating in quadrature an offset carrier with the input signals and adding a bias to the modulated carrier, before the signals are applied to the AODs of the processor.

A different signal processing operation can be performed with the system of Fig. 6, if we set $a(t) = s(t) \exp(jk_1 t^2 + jk_2 t^2)$, $b(t) = \exp(-jk_1 t^2)$, and $c(t) = \exp(-jk_2 t^2)$ [39]. The output in this case takes the following form:

$$d(x, y) = \exp[-jk_1(x/v)^2 - jk_2(y/v)^2] \int s(t) \exp[-j(2k_1 x/v)t] \exp[-j(2k_2 y/v)t] dt. \quad (11)$$

If we neglect the multiplicative quadratic phase term, $d(x, y)$ is the Fourier transform of the signal $s(t)$, calculated via the chirp-z algorithm, in both spatial dimensions. The spectrum is displayed in both dimensions as a function of position with $k_1 x/2\pi v$ and $k_2 y/2\pi v$ being the frequency variables in the x and y directions, respectively. The maximum frequency of $s(t)$ for which the spectrum is calculated in the x and y dimensions is $k_1 A/2\pi v$ and $k_2 A/2\pi v$, respectively, where A is the aperture of both AODs in Fig. 7 (i.e., the maximum value for x and y). The frequency resolution is equal to the maximum frequency divided by the number of pixels in each dimension of the 2-D detector array or the space-bandwidth product of the AODs, whichever is smaller. If k_1 is chosen so that $k_1 A/2\pi v$ is equal to the bandwidth B of $s(t)$, then the entire spectrum of $s(t)$ is displayed in the x direction with relatively low resolution.

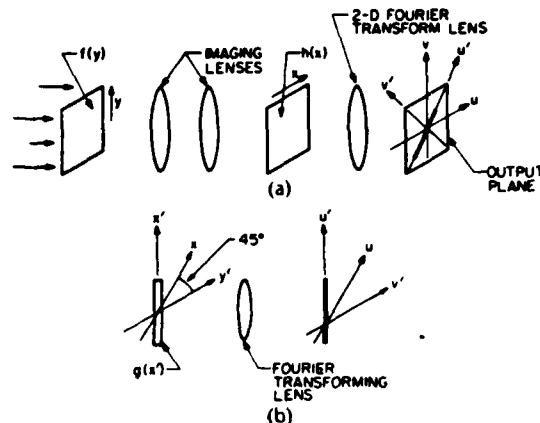


Fig. 7. Triple-correlator architecture.

If k_2 is chosen to be equal to k_1 divided by the number of pixels of the detector (let us take this number to be 10^3), then only $1/10^3$ of the spectrum is displayed in the y direction at each position x . The resolution in the y dimension is then equal to the bandwidth of $s(t)$ divided by 10^6 , assuming that the integration time, T , on the detector is at least $10^6/B$ seconds. The combination of the broad-band, low-resolution transform in x and the high-resolution, low-bandwidth transform in y produces the 2-D "folded" spectrum or, equivalently, the 2-D raster recording of 10^6 samples of the 1-D Fourier transform of the signal $s(t)$. The formation of the 1-D spectrum in two spatial coordinates with this architecture, allows the display of 10^6 spectral samples since 2-D detector arrays of this size are available, whereas the size of 1-D detector arrays is on the order of 10^3 . The triple-product processor can also be used to calculate the "folded" correlation of 1-D signals longer than the acoustic delay of the AODs used in the implementation. This is possible for signals that can be written as the product of two shorter codes [40]. The advantage of the folded correlation over a 1-D implementation is that two very long signals can be correlated and at the same time long differential delays between the two signals are permitted. Either one of these properties can be achieved with a 1-D implementation, but they can be obtained simultaneously only with the 2-D implementation.

The product of two matrices can also be computed by this same architecture using the method described by Athale *et al.* [41]. This is accomplished by applying the columns of the first matrix sequentially to the first AOD in Fig. 6 while the rows of the second matrix are applied to the second AOD. At periodic time instances, when the acoustic signals in the two AODs correspond to one complete column and row of the matrices, the LED is pulsed and the outer product between the two vectors is formed and detected at the output plane of the system. The accumulation on the 2-D detector of the outer products of all the columns of the first matrix with the corresponding rows of the second matrix, results in the formation of the product matrix. In the vector-matrix multipliers we considered earlier (Figs. 3 and 4) both spatial dimensions are used to perform an operation simpler than the product of two matrices. The reason it is possible to perform directly more complex operations with the triple-product processor is that the detector is used to store and accumulate partial results, i.e., data are processed

in the time domain in addition to the two spatial coordinates. Interestingly, however, the triple-product processor that forms the matrix-matrix product is less powerful than the vector-matrix multiplier of Fig. 3. A matrix-matrix product can be performed with the system of Fig. 3 by repetitive vector-matrix multiplications (inner products). The number of inner products that must be computed with this approach is equal to the number of outer products that must be computed with the triple-product processor. The vector-matrix multiplication is computed by the system of Fig. 3 in a time equal to the inverse of the modulation bandwidth of the LEDs, which can be 10 ns or less. The outer product is computed in the system of Fig. 7 in the time it takes to load the AODs, typically 1 μ s or more. Thus the simpler system of Fig. 3 has the speed advantage. This is a consequence of the high degree of parallelism at the input and output stages of this processor, a feature however that in practice is not desirable because of the need to handle electronically many parallel broad-band channels. This tradeoff between speed and parallelism at the input and output stages of the optical system is frequently encountered in the design of optical processing systems.

We now discuss an architecture in which the 2-D space is ingeniously used to perform the correlation of three 1-D signals. The triple correlation is equivalent in the frequency domain to the product of the 1-D transforms of the three signals. All three transforms can be produced with 1-D optical systems (Fourier transforming lenses) but only double products of these transforms are produced in one dimension by detecting their interference. The triple product of the transforms can be formed if both transverse dimensions of the optical system are used, with the processor shown in Fig. 7. This system is a simplified version of an architecture proposed by Vander Lugt [42]. The first function, $f(y)$, is recorded as a 1-D transparency in the vertical direction, and it is imaged through a second horizontal transparency whose transmittance is the second function, $h(x)$. The amplitude of the light following the second transparency is modulated by the product $f(y)h(x)$. A spherical lens forms the 2-D Fourier transform of this product and produces at the output plane of the system in Fig. 7(a) a light distribution whose amplitude is modulated by the product $F(v)H(u)$. F and H are the transforms of f and h , respectively, and v and u are the spatial frequency variables at the output plane. If the (u, v) coordinates are rotated by 45° to obtain the new coordinates $u' = (u + v)/\sqrt{2}$, $v' = (u - v)/\sqrt{2}$, the field at the output plane takes the form $F[(u' - v')/\sqrt{2}]H[(u' + v')/\sqrt{2}]$. At $v' = 0$ the amplitude of the field is modulated along the u' direction by the product of the two transforms. The product of the three transforms can now be formed by detecting the interference of this pattern and the transform of the third signal. The third transform is produced in a separate portion of the system, shown in Fig. 7(b). A 1-D transparency of the third signal, $g(x')$, is recorded in a direction that is at 45° with the (x, y) axes and parallel to the u' axis. The focal length of the Fourier transforming lens in Fig. 7(b) is $\sqrt{2}$ times longer than the one in Fig. 7(a) to compensate for the scaling that occurs from observing the spectra along the diagonal in Fig. 7(a). When the interference of the output light from the two systems in Fig. 7(a) and (b) is detected along the u' axis, one of the terms in the interference pattern consists of the product of the three transforms. The

triple correlation can now be formed by inverse Fourier transforming this term. In Vander Lugt's implementation all three transparencies are implemented with AODs. Therefore, there is a proportional temporal frequency shift at each spatial frequency and this makes it possible to obtain the inverse transform as the temporal signal at the output of a detector that integrates all the light along the u' axis. The particularly interesting feature of this architecture is the use of three AODs oriented in three separate directions. The use of more than two 1-D devices in nonperpendicular directions is the possibility that is only beginning to be explored and will likely result in new 2-D architectures in the future.

The systems that have been discussed here are only a portion of the possible 2-D architectures for processing 1-D signals, but through these examples it becomes apparent that the 2-D space provides not only additional processing power but also versatility. A much broader class of linear operations can be performed on 1-D signals with a 2-D processor and each operation can usually be performed in several different ways. The latitude in the design allows the optimization of the architecture to obtain the most suitable characteristics (input/output format, programmability, processing power) for each application.

IV. PROCESSING OF 2-D SIGNALS

It was stated in the Introduction of this paper that the 2-D processing capability of optics makes this technology particularly well suited for image processing applications. 2-D linear operations such as the Fourier transform and correlation can be directly performed optically if the input image is recorded as a 2-D transparency. When the processor is implemented with 1-D input devices, we are faced with the problem that the entire image cannot be entered simultaneously in the optical processor. Typically, images with 10^6 pixels are processed, whereas the space-bandwidth product of 1-D devices is limited to approximately 10^3 . One possible solution to this problem is the segmentation of the image processing operation into smaller tasks, each one of which is implementable with an optical system that uses 1-D input devices. The partial results can then be stored electronically in order to synthesize the final result when all the subtasks are completed by the optical processor. Such an approach is indeed possible and feasible. In many instances, however, it is not only the processing power that makes the optical implementation attractive, but equally important, its low power requirement, size, and cost. The use of the large, high-speed electronic memory that is needed in this approach can eliminate these advantages. The need for an external memory can be avoided by using the detector of the optical system as the memory in which the partial results are temporarily stored. The partial results can be computed by spatial integration in the optical system and accumulated through temporal integration on the detector to produce the final result. A 2-D space can be constructed, in this manner, from one spatial dimension and time, in which an image can be represented and processed optically by temporal and spatial integration (TSI). Alternatively, the partial results can be computed with a time integrating optical processor, and thus a purely time integrating optical image processor can be implemented. In

either case, the use of the 2-D optical system is essential, since the image must be processed in each of its dimensions along a separate axis of the optical system.

TSI processing was originally proposed by Bader [43] and Psaltis and Casasent [44] for spectrum analysis of 1-D signals. The spectrum of 1-D signals can be optically computed with a 1-D space or time integrating acoustooptic spectrum analyzer. The two implementations have complementary characteristics: the space integrating implementation provides wide bandwidth but relatively low resolution, whereas the time integrating system can produce a high resolution spectrum over a relatively small bandwidth. A high-bandwidth, high-resolution spectrum can be produced with a 2-D optical processor using a combination of temporal and spatial integration. The space integrating spectrum analyzer is shown in Fig. 8(a). If the signal that is

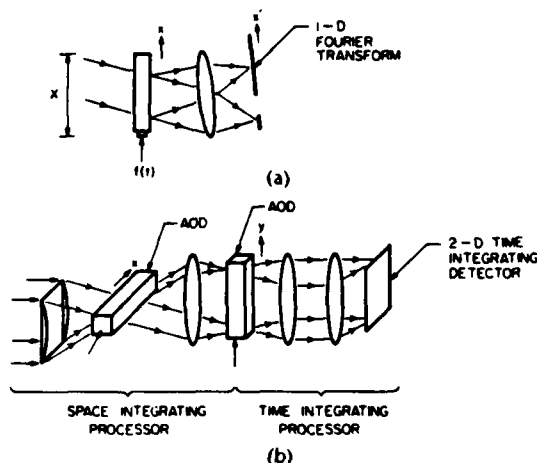


Fig. 8. Time and space integrating spectrum analyzer.

applied to the AOD is denoted by $f(t)$, then the amplitude of the light at the back focal plane of the Fourier transforming lens is modulated by

$$g(x', t) = \int_x f(t - x/v) \exp(j2\pi ux) dx \\ = \{F(vu) * \text{sinc}(Xvu)\} \exp(-j2\pi vut). \quad (12)$$

X in the above equation is the aperture of the AOD, $u = x'/\lambda F$ is the spatial frequency variable, x' is the spatial coordinate at the output plane, F is the focal length of the lens, and $*$ denotes convolution. The spatial modulation at the output is the Fourier transform of $f(t)$, $F(vu)$, smoothed by the convolution with a sinc function due to the finite aperture of the AOD. The frequency resolution is limited by the width of the main lobe of the sinc function and it is equal to v/X hertz. The amplitude of the output light is also temporally modulated sinusoidally at each spatial frequency location at a temporal frequency proportional to the spatial frequency u . If we consider all the light incident on a single spatial frequency resolution cell (with spatial width $\lambda F/X$), then the bandwidth of the temporal modulation of this light is only v/X , equal to the frequency resolution of the space integrating processor. We can therefore use this temporally modulated light as an input to a chirp-z, time integrating processor operating in the orthogonal dimension (Fig. 8(b)) and thus produce a high-res-

olution spectrum of the narrow-band temporal signal localized at each resolution element of the space integrating processor. All the coarse resolution elements can be simultaneously processed by a single multiplexed time integrating spectrum analyzer, as we discussed earlier (Fig. 2). The temporal bandwidth of the light within every one of the spatial frequency resolution cells is the same (v/X), but the center frequency of the temporal modulation at each position is nv/X , where n is an integer that increases in proportion to the position of the resolution cells of the space integrating system. The signals from all cells must be heterodyned to baseband before they can be simultaneously processed by the multichannel time integrating correlator. A pulsed light source with pulse repetition frequency v/X , can be used to accomplish this function. If the individual pulses are sufficiently narrow, the temporal modulation of the light source can be approximated by a series of delta functions

$$\sum_m \delta(t - mX/v) = (v/X) \sum_m \exp[(jmv/X)t]. \quad (13)$$

The amplitude of the light at the output of the space integrating system is modulated by the product of (12) and (13). We see from (13) that the periodic pulse waveform of the light source produces all the harmonics of the temporal frequency v/X . Therefore, at the n th resolution cell, the center frequency nv/X of the modulation will be mixed with the n th harmonic of the source to produce a temporal signal at baseband. Since this happens for all n , there is a baseband signal at each resolution cell. The multichannel time integrating processor produces the fine frequency transform from the baseband signal in each channel while it averages to zero all nonbaseband signals and forms on the 2-D detector array at the output of the time integrating system the folded spectrum of $f(t)$.

TSI processors have been applied to other 1-D signal processing problems such as the computation of the correlation of long 1-D signals and ambiguity functions [45]–[47]. The architecture that produces the folded spectrum of a 1-D signal is of particular interest to this discussion, however, because it suggests a methodology for using TSI to do image processing. The folded spectrum was first discussed in the context of optical signal processing by Thomas [48], who showed that if a long 1-D signal is raster recorded as a 2-D transparency, then the 2-D Fourier transform of this transparency is the raster recording of the transform of the original 1-D signal (i.e., the folded spectrum). Thomas' result can be applied, in reverse, to the TSI processor: the folded spectrum produced by the TSI processor can be interpreted as the 2-D Fourier transform of the image that forms by raster recording the 1-D input signal. Each horizontal line of this image corresponds to a portion of the 1-D input signal, with duration equal to the acoustic delay through the aperture of the AOD. The 2-D Fourier transform of an image can be computed with the system of Fig. 8, by using a TV camera to detect the image and setting the duration of each horizontal video line equal to the acoustic delay of the input AOD. The video from the TV camera is applied as the 1-D input signal to the processor and the 2-D Fourier transform of the image forms on the 2-D detector at the output plane in Fig. 8. Notice that the pulse repetition frequency of the light source of the system in Fig. 8 is equal to the inverse of the acoustic delay. An alternate interpreta-

tion of the image transforming TSI processor is as follows: each line of the input image is scanned by the TV camera and entered in the AOD; the light source is pulsed to read out the current video line which is transformed by spatial integration with the lens; the different lines of the image are sequentially scanned and entered into the optical system where the Fourier transform along the second dimension of the image is produced by temporal integration in the second dimension of the optical processor.

The TSI processing concept can be applied to a more general class of image processing operations. Specifically, it is possible to realize any 2-D linear operation that has a kernel separable in the two dimensions or shift invariant operations. 2-D linear operations with separable, but otherwise arbitrary kernel, can be implemented with the processor shown in Fig. 9 [49]. The input image is denoted

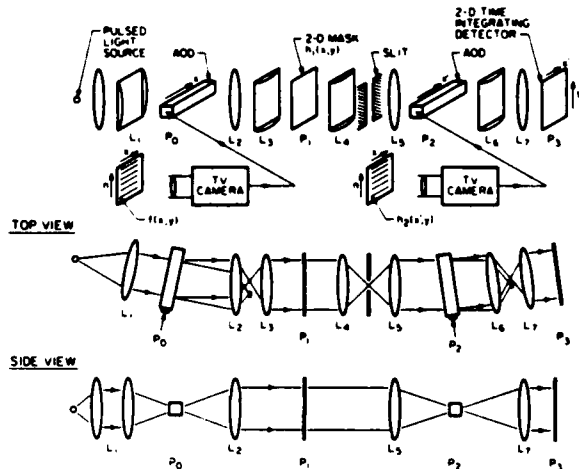


Fig. 9. Time and space integrating image processor, for the implementation of linear operations with space-variant, separable kernel.

by $f(x, n)$, where x is a continuous variable along one of the dimensions and n is an integer that enumerates the lines along the second dimension. A TV camera is used to scan sequentially the lines of the image $f(x, n)$, and the video is applied to the first AOD in Fig. 9. The light source is pulsed periodically to read out separately each line from the AOD. The light diffracted by the AOD is imaged in the horizontal (x) direction and expanded uniformly in the vertical direction to illuminate a mask placed in plane P_1 with transmittance $h_1(x, y)$. The light transmitted through this transparency is imaged in the horizontal direction, by the lenses L_4 and L_5 , onto a second AOD that is oriented horizontally as well. A slit is placed at the focal plane of lens L_4 that blocks high spatial frequency components in the x direction. If the slit is sufficiently narrow (approximately equal to the width of one resolution spot), then the image at P_2 is spatially integrated along the x direction. The second AOD is modulated by the video signal from a second TV camera, which is pointed at an image $h_2(n, x')$. x' is the horizontal coordinate at the plane of the second AOD, and n indicates the line number of the second image. The two cameras operate in synchronism, so that the acoustic signal in the second AOD is the n th line of the image h_2 at the instant the signal in the first AOD is the n th line of

the input image f . The light diffracted by the second AOD is imaged horizontally by lenses L_6 and L_7 to the output plane, where a 2-D detector array is placed. In the vertical direction, the light transmitted through the 2-D transparency, $h_1(x, y)$, is imaged onto the output plane by lenses L_5 and L_7 . The charge that is photogenerated on the detector due to a light pulse that occurs when the n th line of the input image is entered in the optical processor, is proportional to

$$g(y, x', n) = \left\{ \int f(x, n) h_1(x, y) dx \right\} h_2(n, x'). \quad (14)$$

The total charge that is generated and accumulated on the detector for all lines is

$$g(y, x') = \sum_n \int f(x, n) h_1(x, y) h_2(n, x') dx. \quad (15)$$

The operation performed by this processor, as described by (15), is a 2-D linear operation on the image $f(x, n)$ whose kernel is equal to the product of the functions $h_1(x, y)$ and $h_2(n, x')$. The image is processed by continuous spatial integration along one of its dimensions and by the summation of the individual lines through temporal integration on the detector, in the other dimension. Even though any separable-kernel linear operation can be implemented with the system of Fig. 9, an improved TSI architecture can often be designed for a particular separable kernel. For example, the TSI 2-D Fourier transforming system that was discussed earlier in this section, is simpler and therefore preferable to the system of Fig. 9, which can perform the same operation if h_1 and h_2 are chosen appropriately. TSI architectures for synthetic-aperture radar image formation [50], [51] and for the calculation of the moments of an image [52] have been designed by Psaltis and Wagner.

2-D shift-invariant operations are a very important class of operations that can also be implemented with TSI architectures, since such operations are useful in many image processing applications. Let $f(x, n)$ denote an input image, where x is a continuous variable and n is an integer denoting the line number of the image, and let $h(x, n)$ be the reference image. The 2-D correlation between f and h is

$$g(x', m) = \sum_n \int f(x, n) h(x + x', n + m) dx. \quad (16)$$

The integer m in the above equation enumerates the lines of the 2-D correlation, whereas the continuous variable x' is used in the other direction. The 2-D correlation is formed by continuous integration in x and summation over n , since the images are represented by discrete lines in our notation. The summation over n is realized by temporal integration on the detector in the implementation of (16) with a TSI architecture. The 1-D integral in the x direction is the correlation of each line of the input image with all lines of the reference image, i.e., for each value of n the correlation is performed for all possible values of m . This operation is performed in a TSI processor, with a 1-D multichannel space integrating correlator in which the different lines of the reference image are stacked in the dimension orthogonal to x . Therefore, the index m in (16) corresponds to the vertical spatial location at which a line of the correlation forms in the optical system. Notice that in (16) the 1-D correlations of the n th input line must be shifted in the m

direction by n pixels before the summation over n is performed. This shift can be implemented in the optical system with a vertical deflector that positions the 1-D correlations appropriately on the detector or, alternatively, by electronic scanning with a CCD detector array. The CCD implementation will be discussed in more detail in the following paragraph. One possible implementation of a space integrating multichannel correlator is through multiplication in the spatial frequency domain of the transform of each input line with the transforms of all the lines of the reference, which can be stored in the optical system as a 1-D Fourier transform hologram [53]. This implementation combined with a scanning CCD has been recently demonstrated experimentally at Caltech [54].

Images can also be processed with a 2-D time integrating processor, using 1-D input devices. This becomes apparent by considering the triple-product processor (an architecture in which temporal integration is performed in both dimensions) which, according to our discussion in the previous section, can be configured to calculate the folded spectrum of a 1-D signal. Therefore, the analogy between the folded spectrum of a 1-D signal and the 2-D Fourier transform can be applied in this case as well, to conclude that the triple-product processor can also be used for the calculation of the 2-D spectrum of an image [61]. Time integrating processors can also be configured to perform a wider class of image processing operations. We will discuss here a time integrating image correlator, shown in Fig. 10, that was

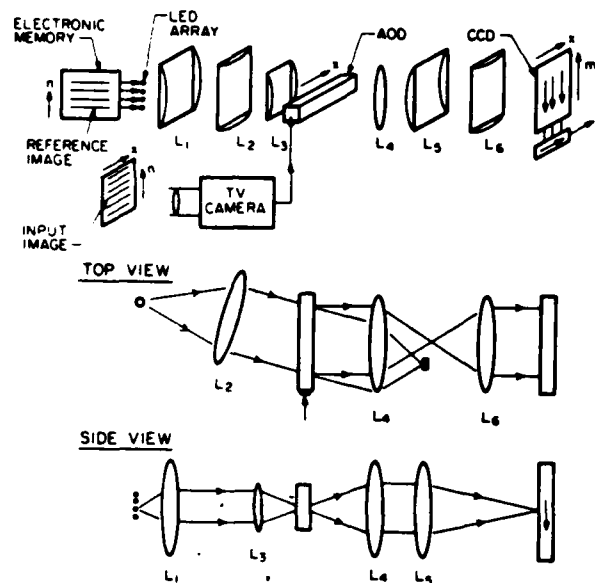


Fig. 10. Time integrating image correlator

recently described by Psaltis [55]. They system consists of a multichannel time integrating correlator in conjunction with a scanning 2-D CCD detector. A TV camera is used to raster scan the input image and the video signal is applied to the AOD. The reference image is stored in an electronic memory (digital or CCD). The memory is interfaced to an array of LEDs in a way that allows each line of the reference image to modulate the intensity of a separate LED. The LEDs are modulated cyclically: during the time one of the lines of the input image is traveling through the AOD, the intensity of the LEDs is temporally modulated by the

lines of the reference image and the LED modulation is repeated every time a new input line enters the AOD. The LEDs illuminate the AOD uniformly in the horizontal (x) direction and the light diffracted by the AOD is imaged horizontally at the output plane of the processor. In the vertical direction, the light distribution at the output plane is an image of the LED array. The charge that accumulates on the 2-D CCD detector array that is placed at the output plane in Fig. 10 during the time that the n th input line travels through the AOD is proportional to

$$g'(x, m, n) = \int f(t - x/v, n) h(t, m) dt. \quad (17)$$

The functions f and h represent the input and reference images, respectively, and the integer m enumerates the lines of the reference image. Since each line of the reference modulates a separate LED and each LED is located at a different vertical position, the integer m in (17) also indicates the vertical position on the CCD detector. The origin of the axis of the time variable, t , in (17) is reset to zero at the beginning of each horizontal scan of the TV camera for notational convenience. At the completion of each horizontal scan, the 2-D charge pattern stored on the CCD is transferred vertically by one pixel. At the initiation of each horizontal scan, the electronic memory also begins to read out the reference image stored in it. When all N lines of the input image have been processed, the signal that was photogenerated by the n th input line, has been transferred vertically on the detector by $(N - n)$ pixels. The total signal that is accumulated on an element of the CCD detector array, located at horizontal position x at the m th row of the array, is given by

$$g(x, m) = \sum_n g'(x, m + n - N, n) \\ = \sum_n \int f(t + x/v, n) h(t, m + n - N) dt. \quad (18)$$

$g(x, m)$ is the correlation between the input image f and the reference image h shifted in the vertical direction by N pixels. The correlation is read out by the CCD in the form of an electronic video signal that can be displayed on a TV monitor or processed further electronically. In this architecture, 1-D devices are used to enter both the input and reference images in the optical processor. Since the reference image is stored electronically, rather than optically on a 2-D SLM, it can be generated with a digital computer and stored in the memory directly. This feature can be very useful in pattern recognition applications, where typically an optimum reference is generated digitally [56] and the need to fabricate a computer-generated hologram is thereby avoided. Furthermore, this image processing architecture has unique flexibility, since the reference image can be periodically reprogrammed by the digital computer.

V. CONCLUSION

The design of optical processing architectures, has, in a sense, gone through a complete cycle. After the early success of 2-D architectures, such as the synthetic-aperture radar processors, 1-D architectures emerged that became practical as 1-D input devices (primarily AODs) were developed. More recently the same 1-D devices have been used in a new generation of 2-D architectures. The primary motivation for the development of these architectures has proba-

bly been the availability of the well-developed 1-D devices. In the process the field has been enriched greatly. A wider class of linear operations is possible with 2-D architectures that are implemented with 1-D devices and more flexibility exists in the way a particular linear operation can be performed. Dynamic processing (time integration and systolic processing) that is feasible with 1-D devices, coupled with the high computational power that is possible in a 2-D optical system, has resulted in the most powerful and flexible optical processors to date. Finally, the abundant processing power that is available in these architectures can be traded for increased accuracy, most notably through the use of the method of binary multiplication by analog convolution [57], [16]. This can lead to optical processors that have accuracy compatible with digital systems and thus increase the applicability of optical computing to more general problems involving linear numerical calculations.

Future architectural designs will certainly be influenced by the state of the art of optical devices. As real-time 2-D spatial light modulators become practical, they will be incorporated in 2-D architectures that use 1-D devices as broad-band input transducers and 2-D devices for massive, programmable storage in the optical system. The development of 2-D input devices that are compatible with dynamic processing, such as multiple-transducer AODs [58], will make new architectural designs feasible. Nonlinear optical devices can also have a major impact; if one- or two-dimensional arrays of optical bistable devices become feasible, then it will be possible to implement optical computers that perform not only linear, but more general computations. Such an optical computer would be more powerful than a digital computer, not because optical gates may be faster than electronic gates, but because the optical computer can be easily configured in the three-dimensional space and thus have massive parallelism and interconnection capability.

ACKNOWLEDGMENT

The author gratefully acknowledges the contributions of Dr. E. G. Paek, K. Wagner, M. Haney, and S. Venkatesh, who are his co-workers at Caltech in the research on acoustooptic image processing.

REFERENCES

- [1] L. J. Cutrona, "Recent developments in coherent optical technology," in *Optical and Electro-optical Information Processing*, J. I. Tippet et al., Eds. Cambridge, MA: MIT Press, 1965.
- [2] E. N. Leith, "Quasi-holographic techniques in the microwave region," *Proc. IEEE*, vol. 59, p. 1305, 1971.
- [3] A. Vander Lugt, "Signal detection by complex spatial filtering," *IEEE Trans. Inform. Theory*, vol. IT-10, no. 2, pp. 139-145, Apr. 1964.
- [4] I. C. Chang, "Acousto-optic devices and applications," *IEEE Trans. Sonics Ultrason.*, vol. SU-23, no. 1, p. 2, 1976.
- [5] T. M. Turpin, "Spectrum analysis using optical processing," *Proc. IEEE*, vol. 69, no. 1, pp. 79-92, 1981.
- [6] D. Casasent, "Spatial light modulators," *Proc. IEEE*, vol. 65, no. 1, p. 143, 1977.
- [7] A. Tanguay, "Spatial light modulators for real time optical processing," in *Future Directions for Optical Information Processing*. Lubbock, TX: Texas Tech Univ., 1981.
- [8] A. A. Bergh, J. A. Copeland, and R. W. Dixon, "Optical sources for fiber transmission systems," *Proc. IEEE*, vol. 68, no. 10, pp. 1240-1247, 1980.
- [9] R. G. Smith, "Photodetectors for fiber transmission systems," *Proc. IEEE*, vol. 68, no. 10, pp. 1247-1253, 1980.
- [10] J. A. Hall, "Arrays and charge-coupled devices," in *Applied Optics and Optical Engineering*, R. R. Shannon and J. C. Wyant Eds. New York: Academic Press, 1980, ch. 8.
- [11] T. Bader and P. Kellman, "Acousto-optic channelized receivers," *Opt. Eng.*, Jan. 1984.
- [12] R. M. Montgomery, "Acousto-optical signal processing system," U.S. Patent 3 634 479, Jan. 1972.
- [13] R. A. Sprague and C. L. Koliopoulos, "Time integrating acousto-optic correlator," *Appl. Opt.*, vol. 15, p. 89, 1976.
- [14] H. J. Caufield, W. T. Rhodes, M. J. Foster, and S. Horvitz, "Optical implementation of systolic array processing," *Opt. Commun.*, vol. 40, p. 86, 1981.
- [15] D. Casasent, J. Jackson, and C. Neuman, "Frequency multiplexed and pipelined iterative optical systolic array processors," *Appl. Opt.*, vol. 22, no. 1, p. 115, 1983.
- [16] P. S. Guilfoyle, "Systolic acousto-optic binary convolver," *Opt. Eng.*, Jan. 1984.
- [17] R. P. Bocker, H. J. Caufield, and K. Bromley, "Rapid unbiased, bipolar incoherent calculator cube," *Appl. Opt.*, vol. 22, no. 6, p. 804, 1983.
- [18] D. L. Hecht, "Multifrequency acousto-optic diffraction," *IEEE Trans. Sonics Ultrason.*, vol. SU-24, p. 7, 1977.
- [19] P. S. Guilfoyle, "Problems in two dimensions," in *Proc. SPIE Conf.*, vol. 341-26, p. 199 (Arlington, VA, May 1982).
- [20] W. T. Rhodes, "Acousto-optic signal processing: Convolution and correlation," *Proc. IEEE*, vol. 69, no. 1, pp. 65-79, 1981.
- [21] D. Casasent and D. Psaltis, "Optical Fourier transform techniques for advanced Fourier spectroscopy systems," *Appl. Opt.*, vol. 19, no. 12, p. 2034, 1980.
- [22] A. V. Oppenheim and R. W. Schaffer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975, p. 321.
- [23] W. T. Tsang, R. A. Logan, and R. P. Salathe, "A densely packed monolithic linear array of GaAs-Al_{0.1}Ga_{0.9}As strip buried heterostructure laser," *Appl. Phys. Lett.*, vol. 34, no. 2, p. 162, 1979.
- [24] J. W. Goodman, A. R. Dias, and L. M. Woody, "Fully parallel, high-speed incoherent optical method for performing discrete Fourier transforms," *Opt. Lett.*, vol. 2, no. 1, p. 1, 1978.
- [25] J. W. Goodman, P. Kellman, and E. W. Hansen, "Linear space-variant optical processing of 1-D signals," *Appl. Opt.*, vol. 16, p. 733, 1977.
- [26] W. T. Rhodes and J. M. Florence, "Frequency variant optical signal analysis," *Appl. Opt.*, vol. 15, p. 3073, 1976.
- [27] R. J. Marks II, J. F. Walkup, M. O. Hagler, and T. F. Krile, "Space-variant processing of 1-D signals," *Appl. Opt.*, vol. 16, p. 739, 1977.
- [28] M. A. Monahan, R. P. Bocker, K. Bromley, and A. Louie, "Incoherent electro-optical processing with CCD's," in *Dig. Int. Optical Computing Conf.*, Apr. 1975 (IEEE Cat. 75 C40941-5C).
- [29] D. Psaltis, D. Casasent, and M. Carlotto, "Iterative, color multiplexed electro-optical processor," *Opt. Lett.*, vol. 4, no. 11, p. 348, 1979.
- [30] R. A. Said and D. C. Cooper, "Crosspath real-time optical correlator and ambiguity-function processor," *Proc. Inst. Elec. Eng. (London)*, vol. 120, p. 423, 1973.
- [31] A. W. Rihaczek, "Principles of high-resolution radars," Mark Resources Inc. Tech. Rep., Marina del Rey, CA, 1977.
- [32] J. D. Cohen, "Ambiguity processor architectures using one-dimensional acousto-optical transducers," *Proc. SPIE*, vol. 180, p. 134, 1979.
- [33] N. J. Berg, M. W. Cassedy, J. J. Abramowitz, and J. N. Lee, "Radar and communication band signal processing using time integration processors," *Proc. SPIE*, vol. 232, p. 101, 1980.
- [34] A. Tarasovich, N. Zepkin, and W. T. Rhodes, "Matrix vector multiplier with time varying single dimensional spatial light modulation," in *Proc. NASA Conf. on "Optical Information Processing"*, (Hampton, VA, 1981, NASA publ. 2207), p. 61.
- [35] P. N. Tamura, J. J. Rebholz, T. Daehlin, and T. Lee, "Real time optical computation of the ambiguity function," *Proc. SPIE*, vol. 241, p. 104, 1980.
- [36] D. Psaltis, and D. Casasent, "General formulation for optical signal processing architectures," *Opt. Eng.*, vol. 19, p. 193, 1980.
- [37] P. Kellman, "Detector integration acousto-optic signal processing," in *Dig. Int. Optical Computing Conf.* (London, England, Sept. 1978), p. 91.

- [38] T. Turpin, "Time integrating optical processor," in *Proc. Soc. of Phot. Instr. Eng. Conf.* (San Diego, CA, Aug. 1978), vol. 154, p. 196.
- [39] P. Kellman, "Time integrating optical signal processing," Ph.D. dissertation, Stanford Univ., Stanford, CA, June 1979.
- [40] D. Casasent and G. Silbershatz, "Product code processing on a triple product processor," *Appl. Opt.*, vol. 21, p. 2076, 1982.
- [41] R. A. Athale, W. C. Collins, and P. D. Stilwell, "High accuracy matrix multiplication with outer product optical processor," *Appl. Opt.*, vol. 22, p. 368, 1983.
- [42] A. Vander Lugt, "Adaptive optical processor," *Appl. Opt.*, vol. 21, no. 22, p. 4005, 1982.
- [43] T. Bader, "Acousto-optic spectrum analysis: A high performance hybrid technique," *Appl. Opt.*, vol. 18, p. 1668, 1979.
- [44] D. Psaltis and D. Casasent, "Time-and-space integrating spectrum analyzer," *Appl. Opt.*, vol. 18, p. 3203, 1979.
- [45] ———, "Spread spectrum time-and-space integrating optical processor," *Appl. Opt.*, vol. 19, p. 1546, 1980.
- [46] T. R. Bader, "Coherent hybrid optical processing," *Proc. SPIE*, vol. 232 (Washington, DC), Apr. 1980.
- [47] G. Silbershatz and D. Casasent, "Hybrid time and space integrating processor for spread spectrum applications," *Appl. Opt.*, vol. 22, p. 2095, 1983.
- [48] C. E. Thomas, "Optical spectrum analysis of large space bandwidth signals," *Appl. Opt.*, vol. 5, p. 1782, 1966.
- [49] D. Psaltis, "Acousto-optics processing of 2-D signals using temporal and spatial integration," AFOSR Rep., Calif. Inst. Technol., Pasadena, CA, May 1983.
- [50] D. Psaltis and K. Wagner, "Real-time optical SAR processor," *Opt. Eng.*, vol. 21, p. 822, 1982.
- [51] D. Psaltis, M. Haney, and K. Wagner, "Real-time synthetic aperture radar processing," in *Proc. NASA Conf. on "Optical Information Processing"* (Hampton, VA, 1983).
- [52] K. Wagner and D. Psaltis, "Real-time computation of moments with acoustooptics," *Proc. SPIE Conf.*, vol. 352-19 (San Diego, CA, 1982).
- [53] D. Psaltis, "Optical image correlation using acousto-optic and charge coupled devices," *Appl. Opt.*, vol. 21, p. 491, 1982.
- [54] D. Psaltis, E. G. Paek, and S. Venkatesh, "Acousto-optic/CCD image processor," in *Proc. Int. Opt. Comp. Conf.* (Cambridge, MA, Apr. 1983, IEEE Cat. 83CH1880-4).
- [55] D. Psaltis, "Incoherent electro-optic image correlator," *Opt. Eng.*, Jan. 1984.
- [56] H. J. Caulfield and W. T. Maloney, "Improved discrimination in optical character recognition," *Appl. Opt.*, vol. 8, p. 2354, 1969.
- [57] D. Psaltis, D. Casasent, D. Neff, and M. Carlotto, "Accurate numerical computation by optical correlation," *Proc. SPIE*, vol. 232-57, p. 151, 1980.
- [58] A. Bardos, "Wideband holographic recorder," *Appl. Opt.*, vol. 13, p. 832, 1974.
- [59] R. V. Johnson *et al.*, "Characteristics of the linear array total internal reflection (TIR) electro-optic spatial light modulator for optical information processing," *Opt. Eng.*, vol. 22, p. 665, 1983.
- [60] R. H. Kingston, B. E. Burke, and K. B. Nichols, "Spatial light modulation using electroabsorption in GaAs CCD," *Appl. Phys. Lett.*, vol. 41, p. 413, 1982.
- [61] W. T. Rhodes, "Space-frequency conversion for image transmission and processing," *Opt. Lett.*, vol. 3, p. 24, 1978.

Optical image correlation using acoustooptic and charge-coupled devices

Demetri Psaltis

An optical processing method is presented which allows the correlation of two images to be computed in real time with a 1-D acoustooptic spatial light modulator as the input device. Two-dimensional processing is accomplished by a combination of spatial and temporal integration. The time-integrating processing is performed by a CCD detector that is operated as an optically addressed correlator.

I. Introduction

Optical image processors normally require at the input stage a real-time 2-D spatial light modulator (SLM)¹ which creates a suitable transparency corresponding to the input image. The relatively immature state of development of 2-D SLMs is a major factor that has prevented the realization of the huge potential of optical image processors. The limitations presently imposed by the SLMs can be overcome either by further developments in device technology or alternatively by modifying the optical system architecture. Several architectural designs have been proposed for this purpose.²⁻⁴ For example, if the optical processor can be implemented with noncoherent illumination the requirements on the optical quality of the SLM can be relaxed, and a self-luminous display (such as a CRT) could be used as the input SLM. In a recent publication by the author³ a 2-D signal processing technique has been presented that does not require the use of real-time 2-D SLMs. This is accomplished by entering the 2-D input data into the optical processor one line at a time using a 1-D acoustooptic SLM.⁵ Each line is filtered spatially by the optical system. The data are processed in the second dimension by integrating the response due to consecutive lines on a time-integrating detector array. Acoustooptic technology is highly developed and consequently high quality acoustooptic SLMs can be fabricated. The use of acoustooptic devices (AOD) rather than 2-D SLMs in the optical processor can lead to improvements in accuracy, speed, and practicality (size, weight, power requirements, cost). In

this paper we describe the application of the time-and-space integrating processing technique to the implementation of an image correlator. Image correlation is a useful signal processing operation in pattern recognition, image restoration and enhancement, and video compression and is also the most powerful operation that can be implemented by an optical computer. Therefore, image correlation is probably the most significant 2-D signal processing operation that can be implemented with the time-and-space integrating method. In the following section we describe the operation of the optical image correlator, and in Sec. III we discuss the unique characteristics of this architecture.

II. Description of the Processor

The time-and-space image correlating architecture is shown in Fig. 1. We will demonstrate that this system computes the cross-correlation $g(x,y)$ of two complex functions $f(x,y)$ and $h(x,y)$:

$$g(x,y) = \iint f(x,y)h^*(x+x,y+y)dx dy. \quad (1)$$

The operation of the correlator in Fig. 1 can be summarized as follows: the optical system is a multichannel 1-D correlator which consecutively correlates each row of the function $f(x,y)$ with all rows of the function $h(x,y)$. The AOD is used to enter the rows of $f(x,y)$ sequentially into the optical system. The 1-D correlations are summed appropriately by operating the CCD detector in the shift-and-add mode.⁶ The key components of the system are the AOD and the CCD detectors.

The AOD consists of a piezoelectric transducer bonded onto an acoustooptic crystal. We denote the voltage applied to the piezoelectric transducer by $s(t)$. The signal $s(t)$ has the form $s(t) = \alpha(t) \cos[\omega_0 t + \phi(t)]$, where the complex envelope $a(t) = \alpha(t) \exp[j\phi(t)]$ is the modulating signal and $\omega_0/2\pi$ is the center frequency of the AOD. The device is illuminated by a collimated light beam incident at the Bragg angle θ_B , where⁷

The author is with California Institute of Technology, Department of Electrical Engineering, Pasadena, California 91125.

Received 7 August 1981.

0003-6935/82/030491-05\$01.00/0.

© 1982 Optical Society of America.

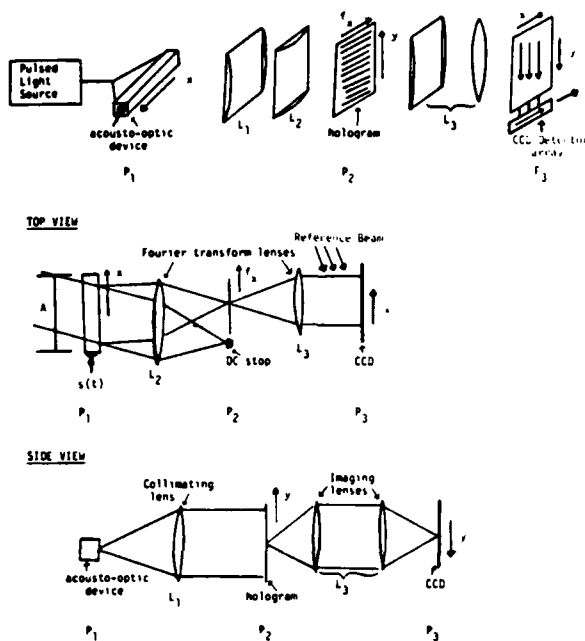


Fig. 1. Image correlator architecture.

$$\sin \theta_B = \frac{\lambda \omega_0}{4\pi v}. \quad (2)$$

λ is the wavelength of light in the acoustooptic crystal, and v is the speed of sound in the crystal. The acoustic wave that is induced by the voltage $s(t)$ causes a portion of the incident light to be diffracted. In the Bragg regime essentially all the diffracted energy is concentrated in the first order. For relatively weak modulation levels the amplitude of the diffracted light $E(x',t)$ is given by⁸

$$\begin{aligned} E(x',t) &\approx c_1 \text{rect} \left(\frac{x' - A/2}{A} \right) a \left(t - \frac{x'}{v} \right) \\ &\times \exp \left[j \phi \left(t - \frac{x'}{v} \right) \right] \exp \left[j \omega_0 \left(t - \frac{x'}{v} \right) \right] \exp \left(j \frac{2\pi \sin \theta_B x'}{\lambda} \right) \\ &= c_1 \text{rect} \left(\frac{x' - A/2}{A} \right) a \left(t - \frac{x'}{v} \right) \\ &\times \exp(j\omega_0 t) \exp \left(-j \frac{2\pi \sin \theta_B x'}{\lambda} \right). \end{aligned} \quad (3)$$

In Eq. (3) c_1 is a constant, x' is along the direction of the acoustic wave propagation, and A is the aperture of the AOD. Thus, the amplitude of the diffracted light is modulated by the complex signal $a(t - x'/v)$ over the spatial window A , Doppler shifted by the traveling acoustic wave by $\omega_0/2\pi$, and deflected around the Bragg angle θ_B . For the rest of our discussion we will not carry the deflection term $\exp[-j(2\pi \sin \theta_B x')/\lambda]$ in the equations since it does not affect the operation of the system we will describe. In Eq. (3) the origin of the x' coordinate is assumed to be at the piezoelectric transducer, and x' increases away from the transducer. In

our analysis it is more convenient to use the coordinate transformation $x = -x' + A$. In the x -coordinate system the origin is at the far end of the acoustooptic crystal, and x increases toward the transducer. Equation (3) can be rewritten as follows:

$$E(x,t) = c_1 \text{rect} \left(\frac{x - A/2}{A} \right) a \left(t + \frac{x}{v} - \frac{A}{v} \right) \exp(j\omega_0 t). \quad (4)$$

We shall use Eq. (4) in the rest of our discussion.

The CCD image sensor consists of a matrix of photosensitive sites where incident photons are converted to electronic carriers which are trapped locally in the depletion region of a MOS junction.⁹ The photogenerated charge packets can be transferred to adjacent pixels along one of the dimensions of the array by applying a clock waveform to the device. When the device is used as an image sensor, each charge packet is shifted continuously along the array after a single exposure to light until it reaches the edge of the device where an output CCD stage transfers the signal to the output pin. In our system the CCD array is exposed periodically. After each exposure the charge that is accumulated in each site is shifted by only 1 pixel. The photogenerated charge due to the most recent exposure is simply added to the charge that is already stored in each pixel. In Fig. 1 the spatial dimension along the shifting direction of the CCD is denoted by \hat{y} . Let the intensity of the illuminating light at the location \hat{y} during the n th exposure be denoted by $I(n,\hat{y})$. The charge generated at the pixel located at \hat{y} during the n th exposure is proportional to $I(n,\hat{y})$. After N exposures ($N \geq n$), the charge $I(n,\hat{y})$ shifts $N - n$ pixels or by a distance $(N - n)\Delta y$ in the \hat{y} direction. Δy is the pixel separation. The charge that is accumulated in the CCD as a function of pixel position \hat{y} after N exposures is given by

$$Q(\hat{y}) = \sum_{n=1}^N I[n,\hat{y} + (n - N)\Delta y], \quad (5)$$

for $\hat{y} = 0$ to $\hat{y} = M\Delta y$, where M is the number of pixels in the CCD and $I(n,\hat{y}) = 0$ for all $\hat{y} \leq 0$. The maximum number of exposures N for which the addition in Eq. (3) can be performed is equal to M . Bromley, and Monahan *et al.*⁶ have used a CCD detector array in this mode of operation to perform a variety of 1-D linear operations. In our system we use the CCD detector to construct an array of parallel 1-D correlators.

In most image recognition applications the input scene and the filter functions are real. For this reason and to present the principles of operation of the system more clearly, we will restrict our analysis initially to real functions f and h . For completeness we will describe the operation of the system with complex signals f and h at the end of this section. The 1-D Fourier transform of a transparency with amplitude transmittance $h(x,y)$ is formed with an astigmatic lens system which transforms in the x direction and images along y .¹⁰ A hologram of the resulting light amplitude distribution is formed by recording its interference with a plane wave reference on photographic film or any other suitable device. The amplitude transmittance of the developed hologram is proportional to¹⁰

$$\begin{aligned}
& |B \exp(-j2\pi \sin\theta_H F f_x) + H(f_x, y)|^2 \\
& = B^2 + |H(f_x, y)|^2 \\
& \quad + BH^*(f_x, y) \exp(-j2\pi \sin\theta_H F f_x) \\
& \quad + BH(f_x, y) \exp(+j2\pi \sin\theta_H F f_x), \quad (6)
\end{aligned}$$

where

$$H(f_x, y) = \int h(x, y) \exp(-j2\pi f_x x) dx,$$

θ_H is the angle of incidence on the hologram of the reference beam, B is its amplitude, F is the focal length of the lens in the x direction, and f_x is the spatial frequency variable which is linearly related to the spatial variable ξ in the plane of the hologram ($f_x = \xi/(\lambda F)$). In Eq. (6) only the term containing H^* contributes to the correlation, and the remaining terms are blocked out in the optical system. Thus the effective complex transmittance of the hologram can be written

$$t_H(f_x, y) = BH^*(f_x, y) \exp(-j2\pi \sin\theta_H F f_x). \quad (7)$$

This hologram is placed in plane P_2 of Fig. 1.

The image $f(x, y)$ is scanned in a raster format to produce a temporal electronic signal $r(t)$. The raster signal $r(t)$ is related to the 2-D function $f(x, y)$ by the following equation:

$$r(t) = f[(t - (n-1)T)/v_s, n\Delta y] \quad (8)$$

for $n = 1$ to N , where v_s is the scanning velocity of the device (such as a TV camera) that produces the raster signal, T is the duration of each raster line [$v_s T$ is the size of $f(x, y)$ in the x direction], and Δy is the resolution of $f(x, y)$ in the y direction [$N\Delta y$ is the size of $f(x, y)$ in the y direction]. In Eq. (8) we assume that $f(x, y) = 0$ for $x > v_s T$ and $x < 0$. $r(t)$ is heterodyned to the center frequency ω_0 and applied to the AOD in plane P_1 of Fig. 1. The modulation $t_A(x, t)$ introduced by the AOD on the amplitude of the diffracted light can be found by substituting $a(t)$ by $r(t)$ in Eq. (4):

$$\begin{aligned}
t_A(x, t) &= c_1 \text{rect}\left(\frac{x - A/2}{A}\right) r\left(t + \frac{x}{v} - \frac{A}{v}\right) \exp(j\omega_0 t) \\
&= c_1 \text{rect}\left(\frac{x - A/2}{A}\right) f\left[\left(t + \frac{x}{v} - \frac{A}{v} - nT + T\right)/v_s, n\Delta y\right] \\
&\quad \times \exp(j\omega_0 t) \quad (9)
\end{aligned}$$

We set $A = v_s T$, i.e., the aperture of the AOD can accommodate exactly one raster line of $f(x, y)$. For convenience we also set $v_s = v$. At time instances $t = nT$ the modulation of the AOD is given by

$$\begin{aligned}
t_A(x, nT) &= c_1 \text{rect}\left(\frac{x - A/2}{vT}\right) f\left[\left(nT + \frac{x}{v} - \frac{A}{v} - nT + T\right)/v_s, n\Delta y\right] \\
&\quad \times \exp(j\omega_0 nT) \\
&= c_1 f(x, n\Delta y) \exp(j\omega_0 nT). \quad (10)
\end{aligned}$$

The rect function can be dropped in Eq. (10) since $f(x, y)$ was defined to be nonzero for $0 < x < vT$. Thus at times $t = nT$, a single line of the function $f(x, y)$ modulates spatially the light diffracted by the AOD. A pulsed light source is used in the system to illuminate the AOD only at the instances $t = nT$. The temporal modulation of the source can be written as

$$t_S(t) = \text{rect}\left(\frac{t - nT}{\tau}\right) \quad (11)$$

for $n = 1$ to N , where τ is the duration of each light pulse, and the pulse shape has been approximated by a rectangular function. The light diffracted by the AOD is modulated by the product of Eq. (9) and (11). We denote this modulation function by $t_{SA}(x, t)$. For $1/\tau$ larger than the bandwidth of $r(t)$, t_{SA} can be approximated by

$$\begin{aligned}
t_{SA}(x, t) &= t_S(t) t_A(x, t) \\
&= \exp(j\omega_0 t) \text{rect}\left(\frac{t - nT}{\tau}\right) f\left[\left(t + \frac{x}{v} - nT\right)/v_s, n\Delta y\right] \\
&\approx f(x, n\Delta y) \exp(j\omega_0 t) \text{rect}\left(\frac{t - nT}{\tau}\right). \quad (12)
\end{aligned}$$

Lens L_1 in Fig. 1 is used to collimate this light distribution in the vertical direction so it illuminates the hologram in plane P_2 uniformly in the y direction. Lens L_2 takes the Fourier transform in the x direction. The amplitude of the light entering plane P_2 is modulated by

$$\begin{aligned}
t_{SA}(f_x, t, n) &= \int t_{SA}(x, t) \exp(-j2\pi f_x x) dx \\
&= F(f_x, n\Delta y) \exp(j\omega_0 t) \text{rect}\left(\frac{t - nT}{\tau}\right), \quad (13)
\end{aligned}$$

where

$$F(f_x, n\Delta y) = \int f(x, n\Delta y) \exp(-j2\pi f_x x) dx.$$

The light immediately after plane P_2 is modulated by the product of Eqs. (7) and (13). The astigmatic lens system L_3 images plane P_2 onto the output plane P_3 in the y direction, while it performs the Fourier transform in the horizontal (x) direction. The amplitude of the light at plane P_3 is given by

$$\begin{aligned}
t_D(\hat{x}, \hat{y}, t, n) &= \int t_{SA}(f_x, t, n) t_H(f_x, \hat{y}) \exp(-j2\pi f_x \hat{x}) df_x \\
&= c_2 \exp(j\omega_0 t) \text{rect}(t - nT) \int F(f_x, n\Delta y) H^*(f_x, \hat{y}) \\
&\quad \times \exp[-j2\pi(\sin\theta_H F + \hat{x}')/f_x df_x] \\
&= c_2 \exp(j\omega_0 t) \text{rect}\left(\frac{t - nT}{\tau}\right) \\
&\quad \times \int f(x, n\Delta y) h(x + \hat{x}, y) dx, \quad (14)
\end{aligned}$$

where $\hat{x} = \hat{x}' + \sin\theta_H F$ and \hat{y} are the horizontal and vertical spatial coordinates, respectively, in P_3 , and c_2 is a constant. The convolution theorem was used to obtain the last form of Eq. (14). The output light distribution is detected interferometrically by the CCD detector array at the output plane to obtain a detected signal proportional to Eq. (14). A reference beam derived from the same source is heterodyned to the center frequency ω_0 of the AOD and made incident on the detector at an angle θ_D . The amplitude of the reference beam is described by

$$t_R(\hat{x}, t) = A \exp(j2\pi \sin\theta_D \hat{x}/\lambda) \exp(j\omega_0 t) \text{rect}\left(\frac{t - nT}{\tau}\right). \quad (15)$$

The signal $I(\hat{x}, y, n)$ that is detected by the CCD is proportional to the time integrated intensity of the sum of Eqs. (14) and (15)

$$\begin{aligned}
I(\hat{x}, \hat{y}, n) &= c_3 \int |t_D + t_R|^2 dt \\
&= \int_{nT-\tau/2}^{nT+\tau/2} |A \exp(j2\pi f_0 x) \\
&\quad + \int f(x, n\Delta y) h(x + \hat{x}, \hat{y}) dx|^2 dt \\
&= \tau c_3 A^2 + \left| \int f(x, n\Delta y) h(x + \hat{x}, \hat{y}) dx \right|^2 \\
&\quad + 2A \left| \int f(x, n\Delta y) h(x + \hat{x}, \hat{y}) dx \right| \cos(2\pi f_0 x), \quad (16)
\end{aligned}$$

where $f_0 = \sin\theta_D/\lambda$. The third term in this equation is the 1-D correlation of the n th row of $f(x, y)$ with all the rows of $h(x, y)$. The correlation forms on the spatial carrier f_0 . Since both f and h are real functions only the amplitude of the fringe pattern is modulated. In the \hat{y} direction the CCD is operated in the shift-and-add mode described earlier. The charge that is accumulated in the CCD after N light pulses can be found by substituting Eq. (16) into Eq. (5):

$$\begin{aligned}
Q(\hat{x}, \hat{y}) &= \sum_{n=1}^N I(\hat{x}, \hat{y} + (n-N)\Delta y) \\
&= c_4 \left\{ NA^2 + \sum_{n=1}^N \left| \int f(x, n\Delta y) h(x + \hat{x}, \hat{y} + (n-N)\Delta y) dx \right|^2 \right. \\
&\quad \left. + 2A \left[\sum_{n=1}^N \int f(x, n\Delta y) h(x + \hat{x}, \hat{y} + n\Delta y - N\Delta y) dx \right] \cos(2\pi f_0 x) \right\}. \quad (17)
\end{aligned}$$

The third term in Eq. (17) forms on a spatial carrier of frequency f_0 . By setting f_0 equal to or larger than the bandwidth of $f(x, y)$, this term can be separated from the other two baseband terms by electronic filtering after the signal from the CCD is converted to a video signal. The envelope of the carrier in Eq. (17) is recognized to be the 2-D correlation described by Eq. (1) with the integration over the continuous variable y replaced by the summation over the discrete variable $n\Delta y$. The correlation pattern is shifted by the constant $N\Delta y$ in the \hat{y} direction. For a CCD detector with N horizontal rows, this means that each slice of the correlation pattern is formed sequentially at the last row of the device. The fast horizontal CCD shift register transfers each line of the 2-D correlation to the output stage of the device where it can be displayed or processed further electronically.

We shall close this section by discussing how the system of Fig. 1 can be used to correlate complex 2-D functions. We assume that we have the complex function $h(x, y)$ recorded either holographically or in a digital memory. The 1-D Fourier transform of this complex function can then be produced and placed in plane P_2 of Fig. 1 either optically or by computer generation. The function $f(x, y)$ is represented by two real functions that correspond to its real and imaginary parts. The two corresponding electronic raster signals are used to modulate in quadrature the carrier that is applied to the AOD. From Eq. (4) we conclude that the modulation introduced by the AOD is proportional to the complex signal that is used to modulate in quadrature the carrier. The optical system performs the complex correlation in the x direction. The phase of the 1-D complex correlation is preserved in the phase of the fringe pattern in Eq. (16). Thus the full 2-D

complex correlation is formed by performing the addition over n in Eq. (17).

III. Discussion

We devote this section to the discussion of the characteristics of the AOD/CCD image correlator. The majority of the potential advantages of this system are consequences of the use of an AOD as the real-time SLM. AODs have been investigated and manufactured for many years. Consequently reliable and relatively inexpensive high quality devices are now commercially available. The simple structure of an AOD relative to most 2-D SLMs allows compact systems to be built with relatively low power requirements and without the need for high voltages. Thus the use of AODs can lead to practically feasible optical image processors in the immediate future. The excellent modulation characteristics of AODs¹¹ (high linear dynamic range, spatial uniformity, and low scatter level) allow the input image to be represented in the optical computer with high fidelity, with a corresponding improvement in overall accuracy. AODs can accept data at rates up to several gigahertz. We do not anticipate however, that this capability can be approached by the AOD/CCD image processor in the near future because of limitations in the speed with which the CCD can be read out. The system also requires a device which produces an electronic raster signal corresponding to the input image. In most applications this device can be a TV camera. The current limitations of CCD and image sensor technologies impose a limit of ≈ 50 -MHz input rate which corresponds to the processing of 200 images, each consisting of 500² pixels in 1 sec. The need for an auxiliary image sensor may be construed as a disadvantage of this technique. Ideally it is preferable to have a device which serves simultaneously as the image sensor and the SLM for the optical computer. In many applications, however, 2-D SLMs are not suitable for direct detection of the live scene, and an auxiliary sensor must be used. The output of the auxiliary sensor is then scanned onto the 2-D SLM either electronically or optically. In such cases, the ability of the AOD/CCD system to accept directly the electronic raster signals is an advantage.

The limitations on the performance of the AOD/CCD processor are imposed primarily by the CCD detector. These limitations are not significantly different from the limitations imposed by the detector on any optical image processor. For continuous processing of frames the input data rate cannot exceed the readout rate in the AOD/CCD processor. The speed at which commercially available CCD detector arrays are normally read out is 5–10 MHz (standard video frame rates). Therefore, the processing speed is limited by the detector if one is restricted to using commercially available devices. CCD detectors which are fabricated specifically for optical signal processing applications can be made with several parallel output stages,¹² which increases the readout speed. For example, a device with 10 parallel readout stages could be read out at 50–100-MHz rates. In addition, current research in GaAs¹³ CCDs shows promise of extending the inherent speed

of these devices to hundreds of megacycles per second. CCD detectors have $\sim 500 \times 500$ pixels,¹⁴ and devices with over one thousand pixels in each dimension are presently under development. The number of pixels of the CCD in the nonshifting dimension determines in this dimension the space-bandwidth product of the images that can be processed. The number of pixels in the shifting direction determines the space-bandwidth product of the filter in the shifting dimension but does not limit the size of the input scene. In other words, this processor is a sliding window correlator in one dimension. This feature can be useful in applications where the input image is sensed by a linear detector array, and the imaging in the second dimension is accomplished by the relative motion between the detector and the object. The dynamic range, noise, linearity, and spatial nonuniformity of the CCD detector are factors that affect the accuracy with which the processed image can be detected. Commercial CCD cameras have dynamic range in excess of 1000:1 with excellent linearity in this range.¹⁴ This range can be extended by designing CCDs with large detector elements that can store a larger photogenerated charge and also by cooling the device and using postdetection electronic processing to minimize detector noise. The detector noise is caused by shot noise of the signal and leakage current, feed-through of clock transients, and thermal noise in the amplifiers.¹⁴ The trade-off between output noise and bandwidth is the major engineering challenge in the design of the AOD/CCD image processor. The portion of the detector dynamic range that is utilized to represent the output signal is proportional to the ratio of the signal to the bias terms in Eq. (17):

$$SBR = \frac{2A \sum_{n=1}^N \int f(x, n\Delta y) h(x + \hat{x}, n\Delta y + \hat{y}) dx}{NA^2 + \sum_{n=1}^N \left| \int f(x, n\Delta y) h[x + \hat{x}, \hat{y} + (n - N)\Delta y] dx \right|^2} \quad (18)$$

For $SBR = 1$, half of the dynamic range is utilized. The SBR however can be significantly lower depending on the functions f and h . It is thus crucial to maximize the dynamic range of the CCD.

We are considering two types of pulsed source for this system: mode-locked gas lasers⁷ and semiconductor lasers.⁷ Gas lasers can satisfy all the requirements of the system (pulse width, peak power, coherence, sta-

bility), but semiconductor lasers are attractive because they are compact, efficient, and relatively inexpensive. The modes in a semiconductor laser, however, are so widely spaced because of the short cavity length that interferometric detection is feasible only with a single-mode laser. Single-mode laser diodes can be fabricated and are even available commercially. However when these diodes are pulsed several modes exhibit gain during the transients. For a well-fabricated diode the transients decay within 1 nsec.¹⁵ Thus, if a pulse with duration of 5 nsec is used, single-mode operation can be obtained 80% of the time.

The operation of the CCD detector in the shift-and-add mode has an important advantageous consequence. Since the signal that eventually reaches the output stage of the detector is a weighted sum of the signal detected by all the detector elements in a single column of the CCD array, the variations in the responsivity among the individual pixels and dark current will average out.¹⁴ In addition, speckle noise which is uncorrelated from pixel to pixel will also average out. Thus the system utilizes coherent light which provides the flexibility of easily synthesizing a filter in the spatial frequency domain but is immune to coherent noise.

References

1. D. Casasent, Proc. IEEE 65, 143 (1977).
2. W. T. Rhodes, Proc. Soc. Photo-Opt. Instrum. Eng. 180, 143 (1979).
3. D. Psaltis, J. Opt. Soc. Am. 71, 198 (1981).
4. M. O. Hagler, R. J. Marks II, E. L. Kral, J. F. Walkup, and T. F. Krile, Appl. Opt. 19, 4253 (1980).
5. J. C. Chang, IEEE Trans. Sonics Ultrason. SU-23, 2 (1976).
6. K. Bromley, M. Monahan *et al.*, Proc. Soc. Photo-Opt. Instrum. Eng. 118, 118 (1977).
7. A. Yariv, *Quantum Electronics* (Wiley, New York, 1967).
8. W. T. Rhodes, Proc. IEEE 69, 65 (1981).
9. D. F. Barbe, Proc. IEEE 63, 38 (1975).
10. J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, New York, 1968).
11. K. Preston, *Coherent Optical Computers* (McGraw-Hill, New York, 1972), p. 163.
12. G. M. Barsbuk, Proc. IEEE 69, 100 (1981).
13. R. C. Eden and I. Deyhimy, Proc. Soc. Photo-Opt. Instrum. Eng. 214, 39 (1979).
14. J. A. Hall, in *Applied Optics and Optical Engineering*, Vol. 8, R. R. Shannon and J. C. Wyant, Eds. (Academic, New York, 1980).
15. K. Lau, "Ultra-high Frequency Dynamics of Semiconductor Injection Lasers," Ph.D. Thesis, California Institute of Technology, Pasadena, 1981.

ACOUSTO-OPTIC/CCD IMAGE PROCESSOR

Demetri Psaltis, Eung Gi Paek and Santosh Venkatesh

California Institute of Technology
Department of Electrical Engineering
Pasadena, CA, 91125

ABSTRACT

Optical image correlators are presented that use acousto-optic and charge-coupled devices as the input and output transducers respectively. Experimental results are presented and the applicability to pattern recognition of a non-linear pseudo-correlation that can also be conveniently computed with these processors, is discussed.

INTRODUCTION

Image processing is perhaps the most natural application of optical information processing because the two dimensions of the optical system and its parallel processing capability are very effectively utilized in this application. The potential of optical image processors however, can only be realized in practice if suitable transducers are available to bring the images into the optical processor with sufficient speed and accuracy and also transducers (detectors) to read-out the processed images. In recent years the development of Acousto-Optic Devices (AOD) and semiconductor detector arrays and light sources, has led to the development of high performance optical processors [1,2]. In this paper we discuss how acousto-optic devices and CCD detectors can be used for optical image correlation. The acousto-optic device is used as the input transducer in the optical processor and it receives the image to be processed in the form of a video electronic signal. The CCD is used as the optical detector in the system and simultaneously as an array of electro-optic correlators. The advantages that can be derived from the use of these relatively mature technologies in the implementation of an optical 2-D correlator are high speed, flexibility, accuracy, small physical size and power efficiency.

In the following section we discuss the general method for performing image correlation with an acousto-optic processor. In section III

we review the holographic implementation of such a processor and the experimental results of this implementations are included in section IV.

GENERAL CONCEPT

The correlation $g(\xi, n)$ of an input image $f(x, y)$ and a reference image $h(x, y)$ is given by:

$$g(\xi, n) = \iint f(x, y) h(x + \xi, y + n) dx dy. \quad (1)$$

We are concerned here with the implementation of this operation with an optical system in which the input image $f(x, y)$ is introduced through an AOD. AOD's are 1-D spatial light modulators with a linear space bandwidth product typically equal to 10^3 pixels. The size of the images that we like to be able to process optically is at least $10^3 \times 10^3$ pixels. Therefore, it is clear that an AOD cannot be used to simultaneously modulate a light beam with an entire image. Typically, a single line of the image can be accommodated by the AOD at one time. This fact dictates our processing strategy: the input image is entered into the optical processor and processed one line at a time. The processed image lines are detected and accumulated on a 2-D CCD camera [2]. The image $f(x, y)$ must be sampled in one of its dimensions (in our notation the y -dimension) so that its lines can be applied sequentially to the AOD. In practice this is accomplished by detecting $f(x, y)$ with a raster scanning TV camera. Since the correlation $g(\xi, n)$ will be computed from the samples of $f(x, y)$ we replace the integration over the continuous variable y in Eq. (1) by a summation:

$$g(\xi, n) = \sum_n \int f(x, n\Delta y) h(x + \xi, n\Delta y + n) dx \quad (2)$$

Δy is the sampling interval in the y direction and if it is chosen to be at the Nyquist rate (assuming f is bandlimited) the correlation g computed via Eq. (2) does not differ from the continuous correlation. N is the number of lines in $f(x, y)$. We can write the shift in the n direction by $n\Delta y$ in the above equation as a

convolution with a delta function $\delta(n+n\Delta y)$ and derive the following form for the correlation $g(\xi, n)$:

$$g(\xi, n) = \sum_n \int \int f(x, n\Delta y) h(x + \xi, n') dx \delta(n' - n - n\Delta y) dn' \quad (3)$$

This form of the correlation function determines directly the implementation algorithm of the acousto-optic image correlator. The term in the brackets in Eq. (3) is the 1-D correlation in the x direction of the n th line of the input image with all the lines of the reference. Several optical implementations of such a multi-channel 1-D correlator are possible. The term in the brackets in Eq. (3) is a function of two variables ξ and n' , and it must be shifted in the n' direction by the distance $n\Delta y$. In other words the 1-D correlations of the first line of the input are shifted by 1 pixel in the n' direction, while the correlations of the N th line are shifted by N pixels. This can be accomplished in an optical system in one of two ways. An optical scanner can be used after the multi-channel correlator to deflect the light by the appropriate distance for each input line. Alternatively (and in most cases preferably) a CCD camera can be used to detect the 1-D correlations and then shift the detected signal by transferring the charge on the CCD. The full 2-D correlation $g(\xi, n)$ is formed by accumulating (summing over n) the shifted 1-D correlation on the time integrating detector array. The operations that must be performed to implement this algorithm optically are summarized in the block diagram of Figure 1.

HOLOGRAPHIC IMPLEMENTATION

We will describe this implementation [3] with the aid of Figure 2. The input pattern is detected by the TV camera. The horizontal sync of the camera is locked to a stable oscillator which acts as the master clock for the entire processor. Each pulse from this oscillator triggers the TV camera to produce an electronic signal corresponding to one of the lines in the input pattern. The video signal is heterodyned to the center frequency of the AOD and after amplification it is applied to the piezoelectric transducer of the AOD. At the end of each horizontal scan by the TV camera, the acoustic wave in the AOD is spatially modulated by the corresponding line of the input pattern. The clock from the master oscillator is delayed by the duration of one raster line and it triggers the pulsed laser that is used as the source in this processor. The duration of the light pulses is made shorter than the inverse of the bandwidth of the video signal. As a result the light diffracted by the AOD each time the laser is pulsed is spatially modulated by one of the lines of the input pattern. The diffracted light enters an astigmatic lens system which expands and collimates the light in the vertical direction. In the horizontal direction the light is Fourier transformed. A 1-D Fourier transform hologram of the reference image $h(x, y)$ in our notation is placed in the Fourier plane of this astigmatic lens system. The hologram contains the reference image transformed in the x direction only. Since the light illuminating the hologram is collimated vertically, the light diffracted by the hologram is modulated by the product of the transform of the current line from the input pattern and all the lines of the reference image. A second

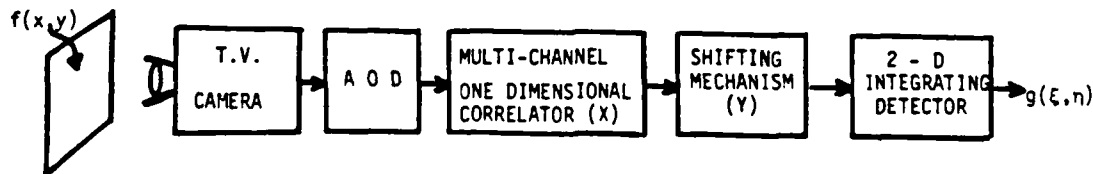


Figure 1. Block diagram of the acousto-optic implementation of the image correlator

A variety of specific optical implementations are possible using this algorithm and the trade-offs involved do not make one of the implementations clearly superior to the rest for all applications. In this paper we discuss a holographic implementation that shares all the characteristics of the algorithm

depicted in Figure 1: the input image is entered in the optical system via the TV camera and the acousto-optic device; a multi-channel 1-D correlator processes each line of the input image and a 2-D time integrating detector is used to accumulate the 1-D correlations and form the full 2-D correlation.

astigmatic lens system is placed after the hologram. The light is now imaged in the vertical direction and transformed again in the horizontal direction. The transform of the product of the transforms produces the correlation (or convolution if desired) between one of the input lines and all the lines of the reference image. The imaging that is performed in the vertical direction causes these 1-D correlations to appear stacked in the vertical direction at the output plane of the optical system. Returning momentarily to Eq. (3), the operation that has been performed up to this point in the processor is the 1-D integral of Eq. (3). A 2-D CCD detector is placed at the output plane of the processor. The photogenerated

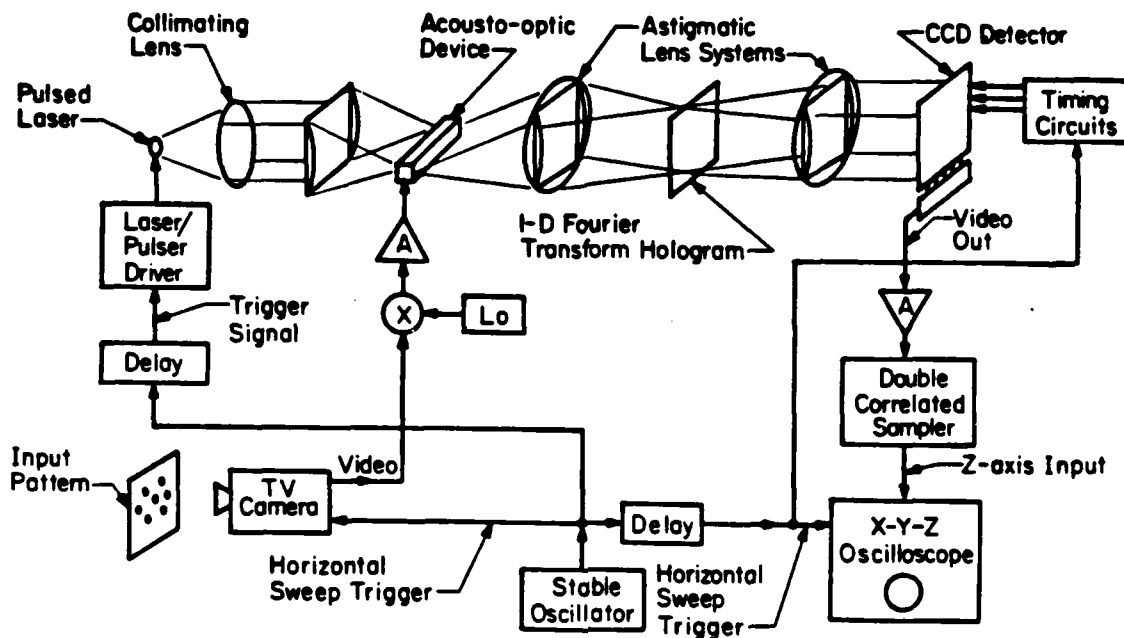


Figure 2. Acousto-optic/CCD image correlator.

2-D charge pattern that is stored in the CCD after it is exposed to the light from each laser pulse, is shifted vertically by one pixel. If we trace the location on the CCD of the charge generated during the n th light pulse we find that after the N th pulse occurs, it has travelled $N-n$ pixels in the vertical direction. To obtain the 2-D correlation we must (according to Eq. (3)) shift the n th line by n pixels. Since N is fixed (it is the total number of lines in the input pattern) the CCD does perform the appropriate shifting and the entire 2-D correlation forms (shifted by N pixels) after integration on the CCD (the summation over n in Eq. 3) of the light incident on it from N light pulses. The correlation output is read-out continuously by the CCD in the form of a video signal which can be displayed or processed further electronically.

In this processor the 1-D correlations that are formed at the output of the optical system modulate the amplitude of the light. This optical wave must be interferometrically detected in order to obtain a detected signal proportional to the field amplitude rather than its intensity. The interferometric detection method [4] and experimental demonstration of it [5] have been discussed elsewhere. In this paper we wish to concentrate

on the 2-D operation that is being performed by the processor of Figure 2, when interferometric detection is not used. In this instance the pattern $g'(\xi, n)$ formed on the CCD is given by the following expression [4]:

$$g'(\xi, n) = \sum_{n=1}^N | \int f(x, n\Delta y) h(x + \xi, n\Delta y + n) dx |^2 \quad (4)$$

The function $g'(\xi, n)$ is similar to the correlation in form, with the important exception that the 1-D correlations in the x -direction are squared before the summation in the second dimension is performed. This operation is clearly non-linear but it is shift invariant. Since $g'(\xi, n)$ can be very conveniently calculated with the acousto-optic image processor, we investigated its properties to determine its possible utility for pattern recognition. Using the Schwarz inequality it can be shown that

$$g'(\xi, n) \leq \sum_n | \int | f(x, n\Delta y) |^2 dx |^2 = G' \quad (5)$$

and the equality holds only for $g'(\xi, n) = G'$. This result is significant because it indicates that $g'(\xi, n)$ is maximized when the input and the reference images are matched and

aligned spatially. Therefore it could be used to recognized patterns. We must however consider the performance of a pattern recognition system based on this operation, when noise is present. A full analysis of this case is beyond the scope of this paper, but we will point out several important facts. First, the signal-to-noise ratio analysis that is typically performed to determine the performance of pattern recognition algorithms that are based on correlation is not appropriate in the nonlinear case. The non-linearity in g' results in noise components at the output that are not normally distributed and therefore the underlying distribution becomes important in the determination of the probabilities of detection and misclassification. Furthermore, patterns that belong to different classes but correlate well (and therefore are likely to be misclassified by a linear correlation algorithm) can be well separated by the non-linear correlation, and vice-versa. For instance, if

$$G = \sum_n \int |f(x, n\Delta y)|^2 dx, \quad (6)$$

is the value of the autocorrelation peak of the image f , we denote by αG the peak value of the crosscorrelation between f and a second image $h(x, y)$ not belonging to the same class as f . α is a constant satisfying $-1 < \alpha < 1$. The non-linear crosscorrelation of the same two functions is denoted by BG' , where G' is defined in Eq. (5) and B is a constant satisfying $0 < B < 1$. It can be shown [6] that given α , B is in the range

$\alpha^2 < B < \alpha$ or $\alpha^2 < B < \alpha$ if the images involved are real and positive. Clearly B can be smaller than α , which intuitively indicates that if this is the case two patterns can be separated more effectively with the non-linear correlation g' . More formally, we calculate the probability of false alarm (the probability of classifying $h(x, y)$ as belonging to the class represented by $f(x, y)$) as a function of the parameter B for the non-linear correlation case. This function is plotted in Figure 3 for an input SNR=1, assuming a linear processing gain of 10^4 , a value of $\alpha = .7$ and for a fixed probability of detection $P_D = .96$. On the same diagram the probability of false alarm that we obtain under the same conditions using linear correlator is indicated. We note that in

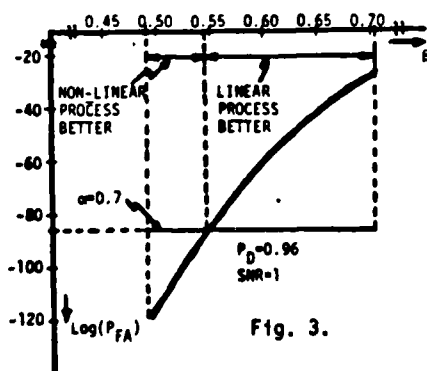


Fig. 3.

Figure 3 there is a very low probability of false alarms for both cases (this is a somewhat superfluous case since α is relatively low and the SNR is high) but we can also see that the probability of false alarm can be smaller for the non-linear correlator. Of course it can also be higher depending on the value of B . B is determined by the images f and h and therefore we conclude from this exercise that a recognition algorithm based on non-linear correlation can out-perform the linear correlator for a certain class of images. In general, the non-linear correlation can provide adequate performance for a large class of pattern recognition applications. A complete statistical characterization of the non-linear correlation will be presented in a future publication.

EXPERIMENTAL PROCESSOR

A photograph of the experimental acousto-optic/CCD image correlator is shown in Figure 4. The apparatus is a hardware implementation of the system shown in Figure 3. The light source used in this experiment is a pulsed laser diode with 50 nsec pulse width. The inverse of the pulse width ($1/50 \text{ nsec} = 20 \text{ MHz}$) must be at least twice the video bandwidth, therefore video signals with up to 10 MHz bandwidth can be processed with the system. The peak power of this laser can be as high as 9W, resulting in an average light power of approximately 10 mW which is adequate for detection by the CCD. A custom fabricated acousto-optic device was the input electronic-to-optical transducer. This device is a shear acoustic wave TeO_2 Bragg cell, with 30 MHz 1dB bandwidth at 820 nm (the laser wavelength), in excess of 400% watt diffraction efficiency and most significantly, the acoustic delay of the device is 70 μsec . The long acoustic delay is required in this processor so that we can accommodate an entire standard video line (63 μsec). The light diffracted by the AOD is demagnified by a factor of approximately 4 with two spherical lenses, in order to match the size of the video lines from the AOD (4 cm) with the size of the transparencies used to fabricate the hologram (1 cm). The demagnified Schlieren image of the AOD is reflected towards the hologram and Fourier transformed. The hologram was formed from a transparency recorded on a high speed holographic plate (to avoid the phase distortion introduced by the plastic substrate of film) with a HeNe laser. The components that are used for recording the hologram are visible in the far end of the optical bench in Figure 4. The holograms are recorded on dichromated gelatin to obtain high diffraction efficiency. We obtain in excess of 20% efficiency with the dichromated gelatin holograms at 820 nm. The light diffracted by the hologram is reflected towards the CCD, imaged vertically and transformed horizontally. The CCD used in the experiment is a commercially available (Fairchild) 1-D device oriented vertically. Special timing and driving electronic circuits were built to operate the device in the time-delay-and-integrate (TDI) mode necessary for this processor. The 1-D array produces slices of the 2-D correlation surface in the n direction, for a fixed ξ value. The slices

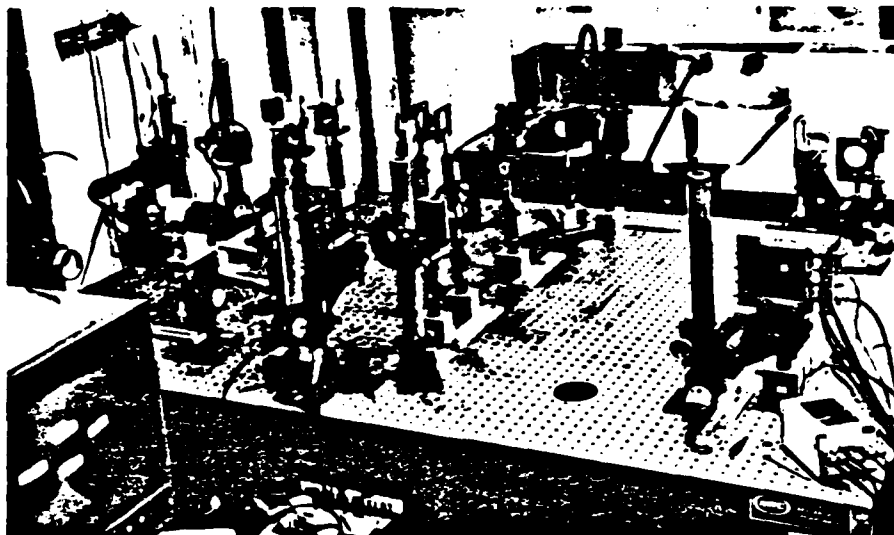


Figure 4. Experimental Acousto-optic/CCD image correlator.

for different ξ values can be observed by translating the CCD horizontally. A 1-D CCD was chosen for this initial experiment, rather than a 2-D CCD which can produce the entire 2-D correlation simultaneously, because its output can be easily monitored on an oscilloscope which has allowed us to easily align and calibrate the system. The input pattern (shown in Figure 5a) was detected by a high resolution TV camera (not shown in

Figure 4). The output signal from the CCD for 5 horizontal positions is shown in Figure 5b. This composite photograph is the 2-D autocorrelation of the pattern in Figure 5a computed by the system in Figure 4. Interferometric detection was not used in these experiments, therefore the non-linear correlation described by Eq. 4 is obtained. The pattern in Figure 5a is in good agreement with the autocorrelation expected from the simple pattern in Figure 5a. We attribute the asymmetries noticeable in Figure 5b to non-linearities and phase distortion in the hologram.

ACKNOWLEDGMENTS

The research reported in this paper is supported by the Army Research Office and the Air Force Office of Scientific Research.

We thank Dr. Peter Kellman for his assistance in the design of the acousto-optic device used in the experiments and Mr. Ed Roos from Crystal Technology for the fabrication of the device.

REFERENCES

- [1] Special issue on Acousto-optic Signal Processing, Proc. IEEE, Vol. 69, 1, Jan. 1981.
- [2] K. Bromley, et. al., Proc. SPIE, Vol. 118, p. 118, 1977
- [3] D. Psaltis, J. Opt. Soc. Am., 71, p. 198, 1981
- [4] D. Psaltis, Appl. Opt., 21, p. 491, 1982
- [5] M. Maney and D. Psaltis, this volume.
- [6] S. Venkatesh and D. Psaltis, to be published.

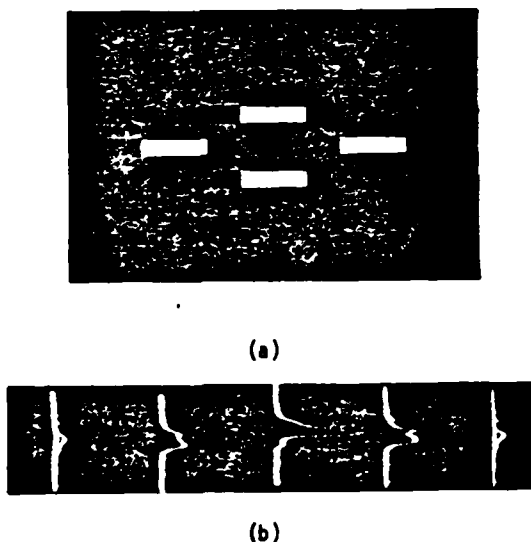


Fig. 5(a). Input Pattern (b) Optically Computed Auto-Correlation

Incoherent electro-optic image correlator

Demetri Psaltis

California Institute of Technology
Department of Electrical Engineering
Pasadena, California 91125

Abstract. A real-time optical image correlator is described. An acousto-optic device and an array of light-emitting diodes are the electronic-to-optical transducers for the input and reference images, respectively, in this architecture. The two-dimensional correlation is formed by temporal integration on a two-dimensional CCD detector.

Keywords: optical computing; optical pattern recognition; incoherent processing.

Optical Engineering 23(1), 012-015 (January/February 1984).

CONTENTS

1. Introduction
2. Description of the architecture
3. Dynamic range considerations
4. Discussion
5. Acknowledgments
6. References

1. INTRODUCTION

Two-dimensional correlation and convolution is one of the most powerful operations that can be performed optically and is useful in a wide range of applications, such as pattern recognition,^{1,2} image restoration/enhancement,³ and synthetic aperture radar.⁴ The classical implementation¹ of an optical image correlator requires 2-D spatial light modulators on which the input image and the Fourier transform hologram of the filter function are recorded. The need for high performance real-time 2-D devices has restricted the applicability of optical image correlators because these devices are generally difficult and expensive to fabricate.^{5,6} Furthermore, the holographic filter is typically recorded on a photographic plate and is therefore fixed. Image correlators can be implemented by multiplication in the spatial frequency domain with incoherent light,⁷ and this can be advantageous because speckle and other coherent artifacts are eliminated while at the same time the need for high quality optical devices is eliminated. The relative disadvantages of this implementation are inflexibility in designing and updating the filter and the strong bias that is formed in addition to the signal at the output of the processor. Rhodes⁸ has introduced a method for image correlation in which the Fourier transform of an image, recorded as a 2-D optical transparency, is produced in the form of a temporal electronic signal at the output of a single detector. The Fourier transform of the filter function can then be stored or produced electronically and multiplied in the time domain with the transform of the input. The correlation is produced by retransforming the product of the temporal signals back to the 2-D space. The primary advantage of this method derives from the fact that the filter is electronically stored or

generated and therefore can be dynamically programmed. The author has recently described^{9,10} an optical architecture in which the input image is entered into the optical processor through an acousto-optic device (AOD) in the form of a video signal produced by a TV camera. The filter is stored in the optical system as a 1-D Fourier transform hologram, and the 2-D correlation is formed on a time-integrating CCD detector array operated in the time-delay-and-integrate (TDI) mode.¹¹ The use of the AOD as the input transducer, which is a well-developed and consequently high performance and practical device,¹² is the most attractive feature of this architecture. The fact that the filter is stored permanently as a holographic transparency restricts its applicability to problems where the filter does not need to be updated dynamically.

In this paper a new architecture is described in which the input transducer is also an AOD, but the filter is electronically stored and is entered into the optical system through an array of light-emitting diodes (LEDs). Therefore, in this architecture the advantages of having a high performance input device are combined with the flexibility of electronic storage of filter function.

2. DESCRIPTION OF THE ARCHITECTURE

A schematic diagram of the architecture is shown in Fig. 1. The input scene is detected by a TV camera, and the video signal is applied to the AOD in Fig. 1. The reference image is stored in a digital or analog (CCD) electronic memory, with multichannel readout capability. The data are stored in memory in a way that allows each line of the reference image to be accessed separately, in parallel. The signal from each channel modulates temporally the corresponding element of the LED array in Fig. 1. If a digital memory is used, a digital-to-analog converter is included in each channel. The memory is read out cyclically in synchronism with the video signal from the TV camera; at the beginning of each horizontal TV scan, the readout of the memory is initiated. The optical system in Fig. 1 is a multichannel time-integrating correlator that forms on the 2-D detector the correlation between each of the signals applied to the LED array and the signal applied to the AOD. In particular, the correlation between the current video line and all the lines of the reference image is formed and stored on the CCD detector. The charge stored on the CCD is transferred vertically by one pixel during the blanking interval of the TV camera. As a result, the correlation between the current video line

Invited Paper OC-104 received Sept. 14, 1983; accepted for publication Sept. 24, 1983; received by Managing Editor Oct. 13, 1983.
© 1984 Society of Photo-Optical Instrumentation Engineers.

and a particular line of the reference image is added to the correlation of the previous video line and the adjacent reference line. The repetitive occurrence of this procedure results in the formation of the full 2-D correlation between the input and reference images, as we will presently show. This architecture is suitable for the correlation of large ($10^3 \times 10^3$ pixels) input images with a relatively smaller reference image ($10^2 \times 10^2$) due to practical considerations which will be discussed further in a subsequent section. In applications such as pattern recognition, this is often the situation we encounter: we are searching for a relatively small object in a large scene.

The input and reference images are denoted by $f(\alpha, \beta)$ and $h(\alpha, \beta)$, respectively, where (α, β) are continuous variables. In this architecture both images are sampled in one of their dimensions before they are processed. In sampled form the images are denoted by $f(\alpha, n)$ and $h(\alpha, m)$, where $0 < n < N$ and $0 < m < M$ are integers that indicate the "line number." N and M are the total number of lines of the input and reference images, respectively. We assume that both functions f and h are real, bipolar with zero mean. f is originally an intensity distribution and therefore positive. When it is converted to an electronic video signal, the dc component can be removed by ac coupling or some other type of filtering. A bias is added to the video signal. The bias is equal to the maximum negative value of f . For convenience all signals will be normalized to unity, and therefore a bias level of one is added. The square root of the biased video signal is formed with an appropriate electronic circuit,¹³ and the signal is heterodyned to the appropriate center frequency and applied to the AOD. For small diffraction efficiencies, the amplitude of the light diffracted by an AOD is proportional to the voltage applied to the device. The intensity of the diffracted light is then proportional to the voltage squared. The square root operation allows a linear mapping of the video signal to light intensity. There are other methods for linear acousto-optic light intensity modulation^{14,15}; however, the method we describe here is the simplest and is well suited to this architecture since we do not require complex signal processing (only bipolar). The intensity modulation introduced by the AOD can be written as follows:

$$I_1(t, \tau) = \text{rect} \left[\frac{\tau - \frac{T}{2}}{T} \right] \times \left[\frac{1}{2} + \frac{1}{2} f(t + \tau, n) \right] \text{rect} \left[\frac{t + \tau - T_1/2}{T_1} \right] \quad (1)$$

where $\tau = x/v$, v is the acoustic velocity in the AOD, x is the horizontal spatial dimension, T is the acoustic delay in the AOD, T_1 is the duration of each horizontal TV scan, and the origin of the x (or τ) axis is taken to be at the transducer of the AOD. The origin of the time axis is reset to zero at the beginning of each horizontal scan. Formally, $t = t' - n(T_1 + T_3)$, where T_3 is the blanking interval (the time that elapses between the end of a horizontal scan and the start of the next one), and t' is the continuous time variable.

The reference image, which is stored in an electronic memory, modulates the array of LEDs in Fig. 1. Each line of the reference image modulates in time a separate LED. The modulation is repetitive; after T_1 seconds have elapsed from the start of each horizontal scan of the TV camera (i.e., at the end of each horizontal scan), the memory is triggered to start reading out in parallel the M lines of the reference. The intensity modulation of the LEDs is given by

$$I_2(t, m) = \text{rect} \left[\frac{t - T_1 - T_2/2}{T_2} \right] \left[\frac{1}{2} + \frac{1}{2} h_2(t - T_1, m) \right] \quad (2)$$

T_2 is the duration of each line of the reference image. Since this processor is designed to correlate relatively small reference images, T_2 will be much smaller than T_1 . The blanking interval T_3 is set equal to T_2 so that the reference is fully read out before the TV camera

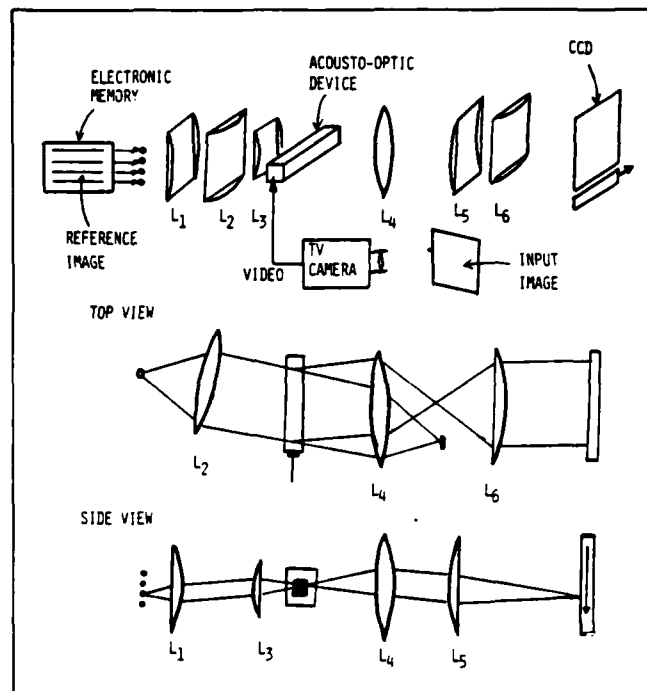


Fig. 1. Incoherent electro-optic image correlator.

starts the next horizontal scan. The index m in Eq. (2) designates the different lines of the reference image. Since there is a one-to-one correspondence between these lines and the LEDs that are stacked in the vertical direction, the index m also designates the position in the vertical direction.

The LED array is imaged onto the AOD in the vertical direction by the combination of the cylindrical lenses L_1 and L_3 . This imaging system has adequate demagnification so that the image of the LED array is smaller than the height of the transducer of the AOD. In the horizontal direction the light emitted from each LED is expanded and collimated by lens L_2 to illuminate the AOD at the Bragg angle. The combination of the spherical lens L_4 and the cylindrical lens L_6 images the AOD onto the output plane, where a 2-D CCD detector is placed. The undiffracted light is blocked at the focal plane of lens L_4 , and only the light diffracted by the AOD reaches the detector. In the vertical direction, lenses L_4 and L_5 also image the AOD onto the output plane. The use of cylindrical optics allows us to set the magnification such that each LED is imaged onto a separate row of the CCD. Again, there is a one-to-one correspondence between the LED array and the rows of the CCD, and therefore the integer m also indicates the row number of the CCD. The charge that is photo-generated on the CCD as a function of horizontal position τ and row number m during the n th scan of the TV camera is proportional to the time-integrated intensity incident on it:

$$C(\tau, m, n) = \int_0^{T_1 + T_3} I_2(t, m) I_1(t, \tau) dt \\ = \text{rect} \left[\frac{\tau - \frac{T}{2}}{T} \right] \int_0^{T_1 + T_3} \text{rect} \left[\frac{t - \frac{T_1}{2} + \tau}{T_1} \right] \text{rect} \left[\frac{t - T_1 - \frac{T_2}{2}}{T_2} \right] \\ \times \left[\frac{1}{2} + \frac{1}{2} f(t + \tau, n) \right] \left[\frac{1}{2} + \frac{1}{2} h_2(t - T_1, m) \right] dt \quad (3)$$

We require that $T \geq T_1 + T_2$, i.e., the aperture of the AOD is sufficiently long to allow the input and reference signals to interact fully. We assume for the rest of this discussion that $T = T_1 + T_2$. Equation (3) can be rewritten in a simpler form by combining all the rect functions with the limits of integration, as follows:

$$C(\tau, m, n) = \int_0^{T_1+T_2} \left[\frac{1}{2} + \frac{1}{2} f(t + \tau, n) \right] \left[\frac{1}{2} + \frac{1}{2} h(t, m) \right] dt$$

$$= \frac{T_2}{4} \left[1 + \frac{1}{T_2} \int_{T_1}^{T_1+T_2} f(t + \tau, n) h(t, m) dt \right] \quad (4)$$

In the last equation we made use of the assumption that $f(t, n)$ and $h(t, m)$ have zero mean. The photogenerated charge pattern is transferred vertically on the CCD by one row after the completion of each horizontal TV scan. The charge that accumulates in the last (M th) row of the CCD after the n th scan is the sum of charges that were generated during the $M - 1$ previous scans in other rows of the CCD in addition to the charge generated during the current scan:

$$C(\tau, n) = \frac{MT_2}{4} \left[1 + \frac{1}{MT_2} \sum_{i=0}^{M-1} \int_{T_1}^{T_1+T_2} f(t + \tau, n - i) h(t, M - i) dt \right]$$

$$= \frac{MT_2}{4} \left[1 + \frac{1}{MT_2} \sum_{n'=n-M+1}^{n'} \int_{T_1}^{T_1+T_2} f(t + \tau, n') h(t, n' + M - n) dt \right] \quad (5)$$

where i is an integer, and $n' = n - i$. $C(\tau, n)$ is the signal that forms at the last row of the CCD after the n th line of the input image has been processed. This signal is transferred to a CCD stage that transfers the charge horizontally for sequential readout. The output signal is the sum of a bias term and the 2-D correlation between f and h . The 2-D correlation is displayed as a function of τ (horizontal position) and the TV line number n . For instance, if a sharp correlation peak is obtained at $\tau = \tau_0$ after the n_0 th line has been processed, we conclude that the input image contains an object matched to the reference h and centered at $\tau = \tau_0$ in the horizontal direction and at the $(n_0 - M)$ TV line.

3. DYNAMIC RANGE CONSIDERATIONS

The architecture described above is implemented with incoherent light. In incoherent systems, bipolar numbers cannot be directly represented since the signals that are processed modulate the light intensity. To overcome this difficulty, the signals are set on a bias. In certain architectures, particularly if complex numbers must be processed, the data are also on temporal and/or spatial carriers.^{16,17} In this case only a bias is necessary since in image processing and pattern recognition both the input and reference functions are generally real. The bias permits bipolar images to be represented, a crucial requirement in pattern recognition where the reference image must be bipolar to provide adequate discrimination between different classes of objects. The insertion of bias at the input stage results in the

formation of a bias on the detector. As a result, the dynamic range of the system, DR' (defined as the ratio of the maximum signal level to the minimum detectable change in the signal), is reduced since a portion of the detector dynamic range, DR , is allotted to the bias:

$$DR' = DR \left(\frac{SBR}{1 + SBR} \right) \quad (6)$$

where SBR is defined as the ratio of the peak value of the output signal to the bias. We assume without loss of generality that the signal attains its maximum value at $\tau = 0$, $n = M$ [see Eq. (5)]. From Eq. (5) we find the SBR to be

$$SBR = \frac{1}{MT_2} \sum_{n'=n-M+1}^{n'=n} \int_{T_1}^{T_1+T_2} f(t, n') h(t, n') dt \quad (7)$$

Since both functions f and h are normalized to unity ($0 < |f| < 1$, $0 < |h| < 1$), the SBR is bounded in the range $0 < SBR < 1$. The upper limit of unity can be achieved only when $|f(t, n') h(t, n')| \approx 1$ for all t and n' . A more typical value is $SBR = 1/2$, which is obtained approximately when h is the combination of a whitening filter followed by a matched filter. The dynamic range of the CCD is approximately $10^3:1$. As an example, we calculate DR' to be equal to 333:1 for $SBR = 1/2$. This dynamic range is adequate for detecting reliably the correlation peak against the noise background. In many applications, however, the input object f is contaminated by additive interference, which we will denote by $g(t, n)$. g can be broadband noise or other objects. We denote by k the ratio of the peak value of g to the peak value of f . If g is a stochastic process, its peak value is taken to be one standard deviation:

$$k = \frac{\max\{g\}}{\max\{f\}} \quad (8)$$

The normalized signal applied to the AOD is now given by

$$\frac{1}{2} + \frac{1}{2(k+1)} [f(t, n) + g(t, n)] \quad (9)$$

Substituting the above in Eqs. (1) through (5), we obtain the following expression for the peak value of the output correlation:

$$C(\tau = 0, n = M) = \frac{MT_2}{4} \times \left[1 + \frac{1}{(k+1)MT_2} \sum_{n'} \int_{T_1}^{T_1+T_2} f(t, n') h(t, n') dt \right.$$

$$\left. + \frac{1}{(k+1)MT_2} \sum_{n'} \int_{T_1}^{T_1+T_2} g(t, n') h(t, n') dt \right] \quad (10)$$

The first term in the above summation is the bias. The second term is the signal, and the maximum value it can attain is $\alpha/(k+1)$, where $0 < |\alpha| < 1$ according to our previous discussion. The third term is the filtered interference signal, and its maximum value is $k/(k+1)$. If k is large, the interference that filters through to the output can, in principle, be approximately equal to one. In pattern recognition, however, the filter h is chosen to enhance only selected objects contained in f , and therefore the interference term is suppressed by a factor approximately equal to the square root of the space-bandwidth product of the filter. Thus, this term is treated as a noise

contribution to the output that reduces the detector dynamic range but does not significantly affect the bias. The system dynamic range in this case is given by

$$DR' = \frac{\alpha DR}{(1 + \sigma_i/\sigma_d)(\alpha + k + 1)} \quad (11)$$

where σ_i and σ_d are the standard deviations of the filtered interference and the detector noise, respectively. In the absence of interference $\sigma_i = 0$ and $k = 0$, and Eq. (11) reduces to Eq. (6). In the presence of additive interference, however, DR' is reduced significantly. Let us consider a specific example. We will take $\alpha = 0.5$ and $DR = 10^3:1$ ($\sigma_d = 10^{-3}$), as before, and we will set $k = 10$ and $\sigma_i = 10^{-2}$. The value for σ_i is calculated by assuming the processing gain of the correlator to be 50. For this case DR' is found to be equal to 3.95:1. This is barely adequate for reliable detection of the correlation peak. This example demonstrates that the bias buildup that unavoidably occurs from the use of incoherent light can result in an unacceptable loss in performance. In most pattern recognition applications, however, the value of $k = 10$ used in the example is not typical. Usually the objects that are encountered are not so deeply buried in the noise. A value of $k = 1$ is closer to what is usually encountered in practice, and this corresponds to a system dynamic range well over 100:1. This dynamic range is sufficient for reliable detection, and therefore the incoherent correlator described in this paper is applicable for pattern recognition problems provided the input signal-to-noise ratio is not exceedingly small.

4. DISCUSSION

The optical correlator originally proposed by Vander Lugt¹ for image recognition is a matched filter. In other words, the reference image is a replica of the object to be recognized. A matched filter is optimum only for the detection of this one object in the presence of white noise. Matched filtering is generally too sensitive to geometrical distortions (scale, rotation) and does not optimally discriminate against objects that do not belong to the class of interest. The performance of an optical pattern recognition system can be improved by enhancing the high spatial frequencies of the reference,¹⁸ using multiple filters,¹⁹ or synthesizing an optimum filter with information collected from a training set of objects.²⁰⁻²² All of these techniques have been previously demonstrated by optical synthesis of appropriate Fourier transform holograms or by computer-generated holography. The electronic storage of the reference image in the architecture presented in this paper provides the flexibility for computing accurately the reference function with a digital computer that is interfaced to the memory in which the reference is stored. Thus, the reference can be updated periodically by the digital computer, and perhaps more importantly, the computer-generated reference can be accurately represented in the optical system. The electronic memory can be either a recirculating CCD analog memory or a digital memory. If a digital memory is used, an array of digital-to-analog converters is incorporated at the interface of the memory to the LED array. LED arrays can be built either monolithically²³ (on a single chip) or by stacking individual LEDs. In either case the size of such an array that is presently feasible is not more than 100 elements. An array of 32 or 64 elements is possible. Laser diode arrays can also be used. Laser diodes have better linearity than LEDs, but the threshold current and the infrared wavelength of laser diodes are relative disadvantages. Several monolithic laser diode arrays have also been fabricated.^{24,25} If a 64 element LED array is used, the reference memory must store 64×64 pixels. For a digital memory with 8 bits of

accuracy in each pixel, this corresponds to a total of 32 kbits of memory. The contents of the memory must be read out at a rate equal to the bandwidth of the video signal. For an input image with 512×512 resolution elements processed at 30 frames per second, this rate is 7.8 MHz or 127 ns per sample. Both digital memories and CCDs can be read out at this rate. The acoustic delay of the AOD must be equal to the duration of each raster line (50 μ s excluding the blanking interval) plus the duration of each line of the reference (5 μ s). TeO_2 devices can, in principle, be fabricated with up to 100 μ s acoustic delay.²⁶ We have tested in our laboratory TeO_2 devices with up to 70 μ s delay. Finally, the CCD must have $512 + 64 = 576$ pixels in the horizontal direction and 64 pixels in the vertical direction. The CCD is read out at standard video rates (7.8 MHz) also. These requirements can be satisfied with CCDs that are available commercially.

The advanced state of the art of the components used in this architecture, which can lead to the implementation of practical (power efficient, compact, inexpensive) processors, and the flexibility provided by the electronic storage of the reference are the most attractive features of this approach. The limited dynamic range that results because incoherent light is used in the implementation and the relatively small size of the reference image that is a consequence of device (LED array) limitations are limitations that are inconsequential for a majority of pattern recognition applications.

5. ACKNOWLEDGMENTS

The research reported in this paper is supported by the Army Research Office and the Air Force Office of Scientific Research.

6. REFERENCES

1. A. Vander Lugt, IEEE Trans. Inf. Theory IT-10, 2 (1964).
2. D. Casasent, Proc. IEEE 67(5), 813 (1979).
3. G. W. Stroke et al., Proc. IEEE 65(1), 39 (1977).
4. E. N. Leith, Proc. IEEE 69(1), 65 (1981).
5. D. Casasent, Proc. IEEE 65(1), 143 (1977).
6. A. R. Tanguay, Jr., Proc. ARO Workshop on Future Directions for Optical Information Processing, Lubbock, Texas, May 1980 (Publisher: Texas Tech Univ.), 52 (1981).
7. J. D. Armitage and A. W. Lohmann, Appl. Opt. 4, 461 (1965).
8. W. T. Rhodes, in *Real-Time Signal Processing II*, T. F. Tao, ed., Proc. SPIE 180, 143 (1979).
9. D. Psaltis, Appl. Opt. 21, 491 (1982).
10. D. Psaltis, E. G. Paek, and S. Venkatesh, in *Proceedings 10th International Optical Computing Conference* (Boston, Mass.), S. Horvitz, ed., Proc. SPIE 422, 204 (1983).
11. K. Bromley et al., in *Optical Signal & Image Processing*, D. Casasent, ed., Proc. SPIE 118, 118 (1977).
12. Special Issue on Acousto-optic Signal Processing, Proc. IEEE 69, (1981).
13. J. D. Graeme, G. E. Tobey, and L. P. Huelsman, eds., *Operational Amplifiers*, McGraw-Hill, New York (1971).
14. R. A. Sprague and C. L. Koliopoulos, Appl. Opt. 15, 89 (1976).
15. P. Kellman, "Time Integrating Optical Signal Processing," Ph.D. Thesis, Stanford University, p. 46 (1979).
16. W. T. Rhodes and A. A. Sawchuck, in *Optical Information Processing*, S. H. Lee, ed., Springer-Verlag, New York (1981).
17. P. Kellman, "Time Integrating Optical Signal Processing," Ph.D. Thesis, Stanford University, Chap. 4 (1979).
18. D. Casasent and A. Furman, Appl. Opt. 17, 1652 (1978).
19. B. D. Guenther, C. R. Christensen, and J. Upatnieks, IEEE J. Quantum Electron. QE-15, 1348 (1979).
20. H. J. Caulfield and W. T. Maloney, Appl. Opt. 8, 2354 (1969).
21. C. F. Hester and D. Casasent, Appl. Opt. 19, 1758 (1980).
22. J. R. Leger and S. H. Lee, J. Opt. Soc. Am. 72(5), 556 (1982).
23. S. Horvuchi, Proc. IEEE Lett. 66(2), 261 (1978).
24. J. D. Crow et al., Appl. Opt. 17(3), 479 (1978).
25. W. T. Tsang et al., Appl. Phys. Lett. 34(2), 162 (1979).
26. I. C. Chang, IEEE Trans. Sonics Ultrason. SU-23, 2 (1976).

Real-time optical synthetic aperture radar (SAR) processor

Demetri Psaltis
Kelvin Wagner

California Institute of Technology
Department of Electrical Engineering
Pasadena, California 91125

Abstract. A real-time optical synthetic aperture radar (SAR) processor is described. The processor utilizes an acousto-optic device as the input electronic-to-optical transducer and a CCD camera that serves as the optical detector and simultaneously performs the focusing of the SAR image in the azimuth direction.

Keywords: two-dimensional signal processing; synthetic aperture radar; acousto-optic devices; imaging; image processing.

Optical Engineering 21(5), 822-828 (September/October 1982).

CONTENTS

1. Introduction
2. Synthetic aperture radar (SAR) signals
3. Acousto-optic/CCD SAR processor
4. Discussion
 - 4.1. Azimuth resolution
 - 4.2. Image size in azimuth
 - 4.3. Range resolution
 - 4.4. Image size in range
 - 4.5. Flexibility
 - 4.6. Dynamic range
5. Experiment
6. Acknowledgments
7. Appendix A: derivation of Eq. (14)
8. Appendix B: derivation of Eq. (15)
9. References

1. INTRODUCTION

Synthetic aperture radar (SAR) still represents the most successful application of optical computing, even though more than 30 years have elapsed since the initial demonstration of the optical SAR processor. Optical computers are used routinely today to form SAR images from radar signals that are collected by aircraft or spacecraft and recorded on photographic film. The requirements of modern radars, however, are often not met by the present film-based optical processors. These requirements include higher image quality (resolution, dynamic range, artifacts, etc.), real-time image formation, flexibility, and on-board processing capability. In order to meet such requirements, digital computers are increasingly being used to form SAR images. The dramatic advances in microelectronics in recent years have made it feasible to construct digital SAR processors that can provide better image quality (due to the highest accuracy and the flexibility in programming a digital computer) than film-based SAR processors. In addition, digital processors can be built (as they have in some cases) that have real-time and on-board processing capabilities. It is not the purpose of this paper to present a comparison of optical versus digital techniques in SAR (the interested reader is referred to Ausherman's paper¹), but it is necessary to point out that a number of problems remain with digital SAR processors. The complexity of these systems results in a very costly, power consuming, relatively large

and heavy processor. Thus, even though it is conceivable to build a low resolution real-time digital processor, for many of the applications this is an impractical solution.

Very significant advances have also been achieved in the area of electro-optics which have resulted in corresponding improvements in the state of the art of optical computers. Specifically, acousto-optic technology has matured, and consequently high quality devices are now available to be used as the input spatial light modulator; semiconductor detector arrays (CCD's and photodiode arrays) have proven to be excellent as detectors in the optical computer, and the semiconductor light sources (LED's and laser diodes) which have been developed primarily for optical communications, are a very efficient light source in many optical processors. These and other developments have made it possible to construct real-time, high performance, power efficient, compact and relatively inexpensive optical computers such as the acousto-optic spectrum analyzer² and the ambiguity function radar processors.³ In this paper we describe an optical SAR processor in which the modern components available for optical computing are fully utilized, and, therefore, we believe that the resulting processor has great potential for applications where on-board, real-time SAR image formation is required.

In Sec. 2 the fundamentals of SAR are reviewed, in order to establish the notation we will be using in the paper. A comprehensive treatment of SAR can be found in Refs. 4 and 5. In Sec. 3 the operation of the optical processor is described in detail, and in Sec. 4 we discuss its performance characteristics. Our experimental results to date are reported in Sec. 5.

2. SAR SIGNALS

The side-looking SAR geometry is depicted in Fig. 1. The vehicle (aircraft or spacecraft) is traveling with a constant velocity v along the η direction and at height h . Coded pulses are emitted periodically through an antenna mounted on the side of the vehicle. In this paper we assume that the pulses are chirp coded. The transmitted waveform $S(t)$ in complex notation is given by

$$S(t) = \sum_n \text{rect} \left[\frac{t - nT}{\tau} \right] \exp[jb(t - nT)^2] \exp(j2\pi v_0 t) \quad (1)$$

where τ is the pulse duration, $1/T$ is the pulse repetition frequency (PRF), b is the chirp rate, and v_0 is the microwave frequency of the radar. The transmitted waveform illuminates a patch of the surface to be imaged (the ground), and part of it is reflected back towards the

Invited Paper TD-103 received Apr. 12, 1982; accepted for publication Apr. 19, 1982; received by Managing Editor May 10, 1982.
© 1982 Society of Photo-Optical Instrumentation Engineers

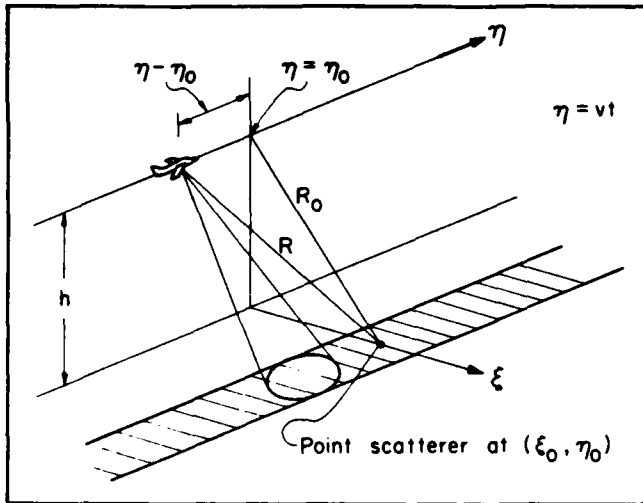


Fig. 1. Synthetic aperture radar geometry.

vehicle and detected by the antenna. The objective of SAR is to form an image of the ground from these reflections. We will discuss how an image of a single point scatterer (impulse) located at coordinates $\xi = \xi_0$ and $\eta = \eta_0$ on the ground can be formed. Since the SAR image formation is a linear operation, knowledge of the impulse response completely describes the system. The signal received by the antenna due to a single point scatterer is

$$r(t) = A(t)S(t - 2R/c) \quad (2)$$

where R is the instantaneous range from the antenna to the point scatterer at (ξ_0, η_0) . $A(t)$ is the far-field pattern of the antenna. From the geometry of Fig. 1, R can be written as

$$R(t) = \sqrt{(\eta_0 - \eta)^2 + R_0^2} = \sqrt{(\eta_0 - vt)^2 + R_0^2} \quad (3)$$

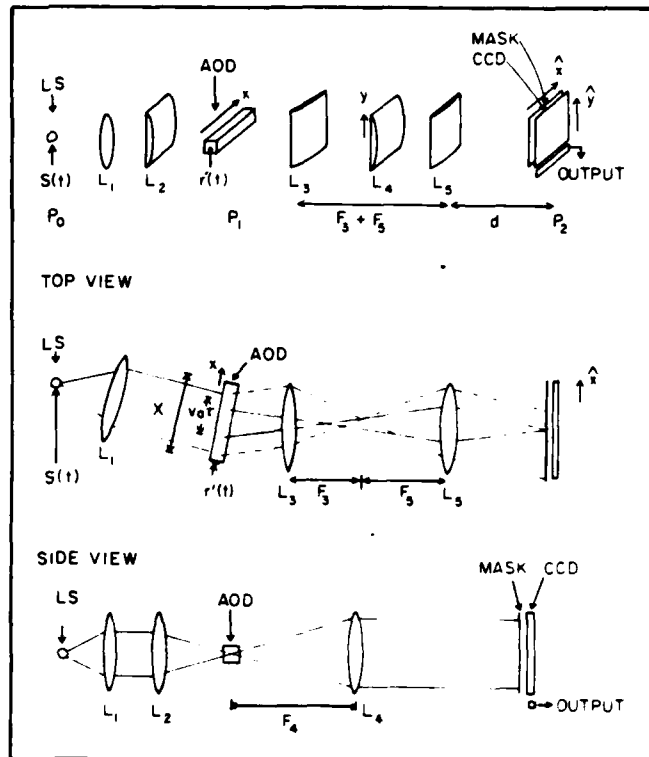
where $R_0 = \sqrt{h^2 + \xi_0^2}$ is the range when the vehicle crosses $\eta = \eta_0$. In most cases, the variation of the range $R(t)$ with time can be neglected within the duration τ of a single radar pulse, and therefore $R(t) \approx R(nT)$. In addition, the range R_0 is typically much greater than the distance $|\eta - \eta_0|$. This allows us to expand the square root in Eq. (3) as follows:

$$R(t) \approx R(nT) \approx R_0 + \frac{(vnT - \eta_0)^2}{2R_0} \quad (4)$$

Using the above equation and Eq. (1), we can rewrite Eq. (2) as

$$r(t) = \sum_n A(nT) \text{rect} \left[\frac{t - 2R_0/c - \Delta\tau - nT}{\tau} \right] \times \exp[jb(t - 2R_0/c - \Delta\tau - nT)^2] \times \exp(j2\pi v_0 t) \exp(-j2\pi v_0 2R_0/c) \exp(-j2\pi v_0 \Delta\tau) \quad (5)$$

The variation of the antenna pattern $A(t)$ during a single pulse can be neglected, and hence $A(t) \approx A(nT)$. The quantity $\Delta\tau = (vnT - \eta_0)^2 / R_0$ has units of time and is very small compared to $2R_0/c$ if $R_0 \gg |\eta - \eta_0|$. Therefore, $\Delta\tau$ can be neglected in Eq. (5) except in the term $\exp(-j2\pi v_0 \Delta\tau)$ because in this term $\Delta\tau$ is multiplied by v_0 which is the very large radio frequency. The term $\exp(-j4\pi v_0 R_0/c)$ is an inconsequential constant phase term and can be omitted. With these modifications, we can write the final form for $r(t)$ as


 Fig. 2. Acousto-optic/CCD synthetic aperture radar processor. (LS—light source, AOD—acousto-optic device, F_1 is the focal length of the lens L_1 , and the broken lines indicate the path of the rays of the reference wave.)

$$r(t) \approx \sum_n A(nT) \text{rect} \left[\frac{t - 2R_0/c - nT}{\tau} \right] \times \exp[jb(t - 2R_0/c - nT)^2] \exp[-j(2\pi v_0/c R_0)(vnT - \eta_0)^2] \times \exp(j2\pi v_0 t) \quad (6)$$

The received signal $r(t)$ must be processed to produce an image of the point scatterer on the ground. In the next section we will describe in detail how the real-time optical processor that is presented in this paper forms the SAR image from the signal $r(t)$. In deriving Eq. (6), a number of simplifying assumptions and approximations were made, which may or may not be justifiable in practice depending on the particular application. We will proceed with the description of the system based on this relatively simple form of the SAR equation [Eq. (6)], and in Sec. 4 we will discuss the necessary modifications to the system for cases in which the conditions set in this section are not met.

3. ACOUSTO-OPTIC/CCD SAR PROCESSOR

A schematic diagram of the processor is shown in Fig. 2. The system is illuminated by a pulsed laser. The light amplitude of the source is modulated by

$$P(t) = \sum_n \text{rect} \left[\frac{t - nT}{\tau_0} \right] \quad (7)$$

where τ_0 is the duration of each light pulse. The spherical lens L_1 collimates the light from the source, and the cylindrical lens L_2 focuses the light in the vertical (y) direction so that it can pass through the aperture of the acousto-optic device (AOD) placed in plane P_1 . The received radar signal $r(t)$ is heterodyned to the center frequency v_1 of the AOD, and a reference signal $B \exp(j2\pi v_2 t)$ is

added to it. The frequency ν_2 is chosen such that the difference $\nu_2 - \nu_1$ is equal to the bandwidth of each radar pulse:

$$\nu_2 - \nu_1 = b\tau/\pi. \quad (8)$$

The need for the reference signal will become apparent shortly. The resulting signal $r'(t)$ is related to $r(t)$ by the following equation:

$$r'(t) = r(t) \exp[-j2\pi(\nu_0 - \nu_1)t] + B \exp(j2\pi\nu_2 t). \quad (9)$$

$r'(t)$ is applied to the piezoelectric transducer of the AOD of Fig. 2. The light diffracted by the AOD is modulated by⁶

$$S_1(t, x) = \text{rect}\left(\frac{x}{X}\right) P(t) r' \left(1 + \frac{x}{v_a}\right), \quad (10)$$

where x is along the direction of propagation of the acoustic wave, X is the aperture of the AOD, and v_a is the speed of sound in the AOD. The undiffracted light is blocked in the focal plane of the cylindrical lens L_3 . The combination of lenses L_3 and L_5 accomplishes two tasks. First, since the two lenses are separated by the sum of their focal lengths, a single plane wave incident on L_3 will be recollimated when it exits the lens L_5 . In addition, the impulse response of the system in the x direction from plane P_1 to plane P_2 is made to be equal to

$$h(x, \hat{x}) = \exp[jb_1(x - \hat{x})^2] \exp(-j2\pi\nu_1 x/v_a). \quad (11)$$

where \hat{x} is the horizontal coordinate in plane P_2 . The constant b_1 can be set by appropriately choosing the focal lengths of L_3 and L_5 and the distance d . The term $\exp(-j2\pi\nu_1 x/v_a)$ reflects the fact that the optical system following the Bragg cell is tilted so that its optical axis coincides with a wave diffracted at the Bragg angle. In Eq. (11) we assume that the effects of the lens apertures are negligible. In the vertical direction y , the light is recollimated by the lens L_4 , and therefore the amplitude of light entering plane P_2 does not vary along y . At P_2 a transparency is placed immediately in front of the CCD camera. The intensity transmittance $T(\hat{x}, \hat{y})$ of this mask is

$$T(\hat{x}, \hat{y}) = \frac{1}{2} + \frac{1}{2} \cos \left[2\pi u_0 \hat{x} + \frac{b_2 \hat{y}^2}{\hat{x}} \right], \quad (12)$$

where u_0 and b_2 are constants that will be specified later. The instantaneous intensity incident on the CCD is given by

$$I(\hat{x}, \hat{y}, t) = T(\hat{x}, \hat{y}) \left| \int S_1(t, x) h(x, \hat{x}) dx \right|^2. \quad (13)$$

The CCD is operated in the shift-and-add mode.⁷ In this mode of operation the device is exposed to light periodically. The photogenerated charge is accumulated on the detector for a short time interval, and after each exposure the entire photogenerated charge pattern is electronically shifted by one pixel along one of the dimensions of the CCD. The photogenerated charge due to the next exposure is simply added to the charge that is already stored in each pixel, and then the process is repeated. In our system the charge transferring is done in synchronism with the PRF of the radar, and the integration period is set equal to the period T of the radar. The charge generated on the CCD during the n th radar pulse is given by

$$\begin{aligned} Q(\hat{x}, \hat{y}, n) &= \int_{t-nT}^{t+nT} I(\hat{x}, \hat{y}, t) dt \\ &= 2Bv_a \tau \tau_0 T(\hat{x}, \hat{y}) A(nT) \text{sinc} \left[\frac{b\tau}{\pi v_a} \left(\hat{x} - \frac{2R_0 v_a}{c} \right) \right] \end{aligned}$$

$$\begin{aligned} &\times \cos \left[\frac{2\pi(\nu_2 - \nu_1)\hat{x}}{v_a} + \frac{2\pi\nu_0}{cR_0} (v_n T - \eta_0)^2 + \phi \right] \\ &+ \text{bias terms} \end{aligned} \quad (14)$$

The derivation of Eq. (14) is straightforward but tedious, and it is presented in Appendix A. This charge pattern gets shifted along the \hat{y} direction by $(N - n)$ pixels on the CCD, after N exposures have taken place. The total charge that accumulates at each pixel located at coordinates (\hat{x}, \hat{y}) after N exposures is

$$\begin{aligned} Q(\hat{x}, \hat{y}) &= \sum_n Q[\hat{x}, \hat{y} + (N - n)\Delta y, n] \\ &= \frac{Bv_a \tau \tau_0}{2} \text{sinc} \left[\frac{b\tau}{\pi v_a} \left(\hat{x} - \frac{2R_0 v_a}{c} \right) \right] \text{sinc} \left[\frac{2Nv_0 v^2 T^2}{c\Delta y R_0} \right. \\ &\quad \left. \times \left(\hat{y} - \eta_0 \frac{\Delta y}{vT} + N\Delta y \right) \right] \cos [4\pi u_0 \hat{x} + \phi] \\ &+ \text{bias terms} \end{aligned} \quad (15)$$

In the above equation Δy is the pixel separation in the CCD, $u_0 = (\nu_2 - \nu_1)/v_a$, and ϕ is a constant phase term. $Q(\hat{x}, \hat{y})$ is the final output of the processor, and its form demonstrates the imaging capability of the system. In the \hat{x} direction, $Q(\hat{x}, \hat{y})$ is a sinc function with width $\pi v_a/b\tau$ and centered at $\hat{x} = 2R_0 v_a/c$. In other words, we obtain the image of the point scatterer located at $\xi = \xi_0$ on the ground, focused in the ξ direction. The resolution in the ξ direction is equal to $\delta_\xi = \pi c/2b\tau$. This is a well-known result for range resolution in radar systems.⁸ The carrier at spatial frequency $2u_0$ arises from the inclusion of the reference signal in the acousto-optic device which allowed us to record on the CCD the phase of the detected signals. This is necessary in this system because the Doppler information that is essential for the focusing in the azimuth (η) direction is encoded in the phase of the range compressed signal $Q(\hat{x}, \hat{y}, n)$. In addition, since the output forms on a carrier, it can be easily separated from the bias term in Eq. (15).

In the \hat{y} direction, $Q(\hat{x}, \hat{y})$ is also a sinc function whose position is proportional to η_0 , the location of the point scatterer on the ground in the azimuth direction. The width of this sinc function is equal to $c\Delta y R_0/Nv_0 v^2 T^2$, and it determines the resolution in azimuth. In ground coordinates η , the width of the sinc function corresponds to an azimuth resolution

$$\delta_\eta = \frac{cR_0}{2Nv_0 vT} = \lambda_0 R_0/2D_s.$$

$D_s = NvT$ is the distance that the vehicle collecting the SAR data travels during the time interval NT , and it is equal to the synthetic aperture of the system. Therefore, the azimuth resolution obtained with this processor conforms with previously derived results⁸ for SAR systems. The entire pattern in shifted in the \hat{y} dimension by the distance $N\Delta y$. For a CCD with N pixels in the horizontal direction, this implies that after the signal is integrated on the CCD for N pulses to produce a focused image, it arrives at the edge of the device (at $\hat{y} = N\Delta y$) where a separate CCD stage transfers an entire line of the data (a slice of the image for each azimuth position) to the output pin of the CCD device. The azimuth slices are continuously produced as long as the flight continues, producing in real time an image of a long strip on the ground parallel to the direction of the flight.

4. DISCUSSION

In this section we discuss some of the issues relative to the performance of the optical processor we have described. We will address

several performance criteria separately and comment on the effects of the processor architecture and the devices used in the system. A complete in-depth analysis of the performance is beyond the scope of this paper.

4.1. Azimuth resolution

In the previous section we derived the formula for the resolution that can be obtained in the along track (azimuth) direction:

$$\delta_\eta = \frac{cR_0}{2Nv_0NT} \quad (16)$$

N is the only parameter in the above equation that is related to the processor. N is the number of radar pulses that are used to form the SAR image of each point scatterer on the ground. In the AO/CCD processor, N is also equal to the number of pixels of the CCD in the shifting direction (\hat{y}). Commercially available CCD cameras have approximately 500 pixels. Several devices with approximately twice this number have recently been developed⁸ or are currently under development. If a CCD with 10^3 pixels is used, the azimuth resolution obtainable with this processor is 22.5 meters on the ground. This figure was obtained by using the parameters of the SEASAT radar⁹ that was flown by NASA. The azimuth resolution will, of course, be different for different radars. To obtain lower resolution, larger CCD's would have to be used. It is unlikely that a significant increase in the size of monolithic CCD cameras will take place in the near future. It is possible, however, to construct larger arrays by interfacing a number of separate chips. As N is increased and finer resolution is achieved, other factors in the processor become limiting, such as optical aberrations, mechanical/electronic stability, and light source coherence over the time period NT . Finally, for low resolution imaging, some of the approximations that were made in Sec. 2 may not be valid. For instance, if the approximation that the range of each point target remains constant as the plane flies past it is not valid, an effect known as range curvature⁴ results. Range curvature can be corrected in our processor simply by tilting one of the cylindrical lenses in the system which introduces barrel distortion that can exactly cancel out the range curvature.

4.2. Image size in azimuth

The AO/CCD processor continuously produces slices of the SAR images in the azimuth direction, and therefore the size of the SAR image is not limited by the processor in this direction.

4.3. Range resolution

The resolution in the range direction (ξ) on the ground is given by

$$\delta_\xi = \frac{\pi c}{2b\tau} \quad (17)$$

The bandwidth of the chirped radar pulses is equal to $b\tau/\pi$, and therefore the resolution in range is limited by the temporal bandwidth of the processor. The bandwidth of the AOD used in the system is not the primarily limiting factor since devices with several gigahertz bandwidth are available. More importantly, the duration of the light pulse τ_0 must be shorter than the inverse of the bandwidth of the signal to avoid smearing of the range compressed signal [Eq. (14)]. For typical radar bandwidths, τ_0 should be several nanoseconds. Several laser sources are capable of producing nanosecond pulses with adequate peak power. We are primarily considering laser diodes because they are compact and power efficient. Laser diodes can lase in a single mode under cw operation (and therefore be coherent), but when they are pulsed, several modes exhibit gain during the transients. The need for coherence in the system imposes with present state-of-the-art devices a minimum pulse width of ~ 20 ns. Therefore, if a laser diode is used in the system, the bandwidth of the radar must be equal to 50 MHz, which corresponds to range resolution equal to 3 m. This can be lowered by an order of magni-

tude if a pulsed gas laser is used. The additional factors mentioned in our discussion on azimuth resolution also become increasingly important for finer range resolution.

4.4. Image size in range

The number of pixels in the range direction is limited by the space-bandwidth product of the AOD or the CCD in the nonshifting direction, whichever is smaller. At the present time, the CCD imposes this limit at approximately 10^3 pixels. The range coverage can be extended by using several processors operating in parallel.

4.5. Flexibility

A potential application for the AO/CCD processor is on-board real-time image formation. In such applications the radar geometry can be expected to change from time to time or even continuously. Enough flexibility must be built into the system so that it can be adaptable to the changing conditions. One method for achieving this is by the use of a real-time 2-D spatial light modulator⁹ in place of the fixed mask in Fig. 2. The impulse response of the processor can then be altered to track changes in the radar geometry.

4.6. Dynamic range

The dynamic range is defined as the maximum unsaturated brightness in the SAR image divided by the minimum detectable level above the noise. In general, a large dynamic range can be observed in a SAR image because the noise detected at radio frequencies is very effectively suppressed through the pulse compression in both range and azimuth that is performed to bring the SAR image into focus. The potential for high dynamic range, however, can only be realized if the processor that forms the SAR image can provide it. In an optical processor, essentially every component of the system contributes to diminish the dynamic range available at the output by adding noise or by introducing nonlinearities and aberrations in the system. In the AO/CCD processor, the CCD detector is the dominant factor that limits the dynamic range. The maximum brightness of the SAR image is equal to the maximum charge that can be stored in each pixel of the CCD, minus the charge that corresponds to the bias terms in Eq. (15) and the dark current that accumulates during the integration time. The bias is signal dependent and is largest at portions of the SAR image that have maximum average intensity. Therefore, relatively sparse objects can be imaged with a larger dynamic range. For a single point scatterer on the ground, the bias can be only 1/2 of the full dynamic range of the CCD, but for a distributed target the bias increases as \sqrt{M} , where M is the average number of point scatterers on the ground that are located at the same range in a distance NvT along the azimuth direction. The minimum detectable level at the output of the processor is determined by the fixed pattern noise of the CCD (the variation of the dark current from pixel to pixel), optical scattering, quantum noise primarily due to the bias, and thermal and amplifier noise from the electronics.¹⁰ Our calculations¹¹ show that the dynamic range of the images that will be produced by the experimental processor we are presently constructing is $\sim 200:1$. A commercially available CCD camera is used in the experimental processor, and post-detection signal processing is applied in order to minimize noise. The dynamic range can be increased to over $10^3:1$ by cooling the CCD to eliminate the dark current and by using a CCD specifically designed for this application.

5. EXPERIMENT

The CCD camera is utilized in this system not only as the optical detector but also as a multichannel correlator by operating it in the shift-and-add mode described earlier. The signal that comes out at each column (at position \hat{x}) of the device is the correlation between the transmittance of the mask as a function of \hat{y} at the same position \hat{x} and the signal corresponding to the temporal variation of the incident light intensity. We have performed an experiment to demonstrate the operation of the CCD as an optical multichannel correlator. The experimental setup is shown in Fig. 3. A light emitting diode (LED) is used as the light source. The intensity of the LED is

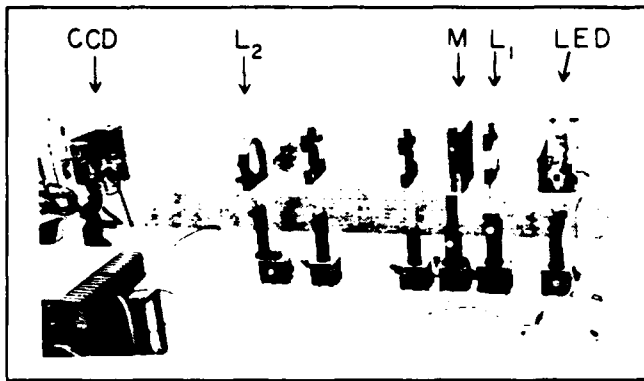


Fig. 3. Experimental electro-optic ambiguity function processor. (LED—light emitting diode, L_1 —collimating lens, M —mask, L_2 —imaging lens, CCD—camera.)

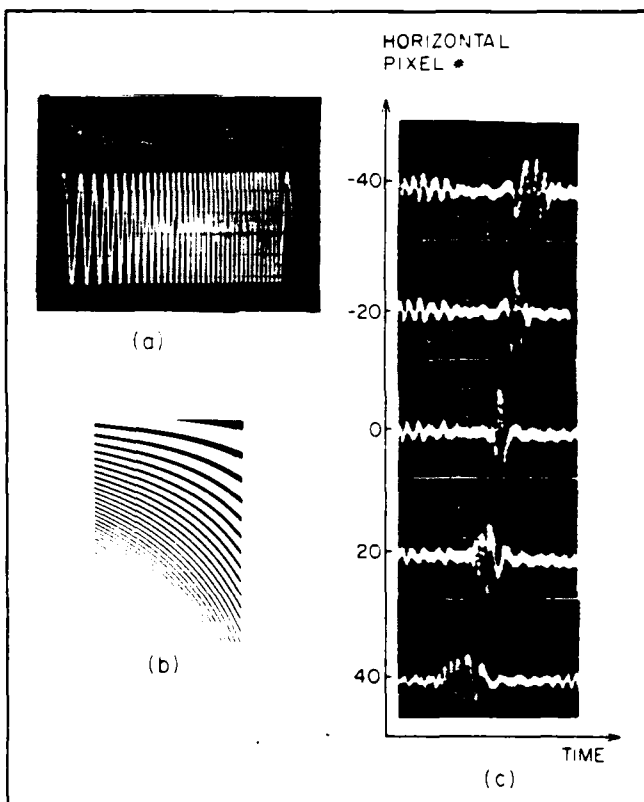


Fig. 4. Ambiguity processor signals: (a) input (LED) waveform; (b) mask; (c) ambiguity function.

temporally modulated by the chirp signal shown in Fig. 4(a). The chirp signal that modulates the LED represents the variation of the intensity detected by the CCD [the function $Q(\hat{x}, \hat{y}, n)$ in Eq. (14)] as a function of pulse number n . The duration of this waveform is equal to NT , and its bandwidth is equal to $v_0^2 NT / cR_0^2$. The bandwidth of the waveform in Fig. 4(a) is 30 Hz, and its duration is 1 s. The CCD we used in this experiment has $N = 512$ pixels in the vertical dimension, and therefore the duration of one second corresponds to a radar pulse repetition frequency equal to $1/T = 0.5$ kHz. These parameters were chosen to be approximately equal to the parameters of an experimental aircraft radar which is flown periodically by the Jet Propulsion Laboratory. We intend to use data collected by this radar to test the processor when it is completed.

The light from the LED is collimated by lens L_1 in Fig. 3, and it

illuminates the mask denoted by M in Fig. 3. A photograph of this mask is shown in Fig. 4(b). The intensity transmittance in the vertical direction is a chirp signal whose chirp rate is proportional to the horizontal position. The signal at the center of the mask matches the waveform that is used to modulate the LED. The mask is imaged by lens L_2 onto the CCD, which is operated in the shift-and-add mode. The output signal from the CCD at each horizontal position is the correlation of the waveform modulating the LED and the transmittance of the mask at that horizontal position. The output waveforms from the CCD from selected vertical locations are shown in Fig. 4(c). This 2-D waveform is known as the ambiguity function¹² of the chirp signal, and Fig. 4(c) displays the familiar tilted-ridge structure of the ambiguity function for this particular waveform.

6. ACKNOWLEDGMENTS

We thank Tom Bicknell from the Jet Propulsion Laboratory (JPL) for the technical information he has provided to us on SAR systems and for loaning to us some of the equipment used in the experiments. The research reported in this paper is supported by NASA and JPL and in part by the Air Force Office of Scientific Research.

7. APPENDIX A: DERIVATION OF EQUATION (14)

$$Q(\hat{x}, \hat{y}, n) = \int_{t-nT}^{t+nT} I(\hat{x}, \hat{y}, t) dt \quad (14)$$

Substituting Eq. (13) in Eq. (14):

$$Q(\hat{x}, \hat{y}, n) = T(\hat{x}, \hat{y}) \int_{t-nT}^{t+nT} \left| \int S_1(t, x) h(x, \hat{x}) dx \right|^2 dt \quad (A1)$$

We will first evaluate the integral over x in the above expression. Using Eqs. (10) and (11):

$$\begin{aligned} g(t, \hat{x}) &= \int S_1(t, x) h(x, \hat{x}) dx \\ &= \int_{\text{rect}} \left(\frac{x}{X} \right) P(t) r' \left(t + \frac{x}{v_a} \right) \exp[jb_1(x - \hat{x})^2] \\ &\quad \times \exp \left(-j \frac{2\pi v_1 x}{v_a} \right) dx \end{aligned} \quad (A2)$$

Substituting the expressions for $P(t)$ and $r'(t)$ from Eqs. (7) and (9):

$$\begin{aligned} g(t, \hat{x}) &= \sum_n \text{rect} \left(\frac{t-nT}{\tau_0} \right) \int_{\text{rect}} \left(\frac{x}{X} \right) \\ &\quad \times \left\{ r \left(t + \frac{x}{v_a} \right) \exp \left[-j2\pi(v_0 - v_1) \left(t + \frac{x}{v_a} \right) \right] \right. \\ &\quad \left. + B \exp \left[j2\pi v_2 \left(t + \frac{x}{v_a} \right) \right] \right\} \\ &\quad \times \exp[jb_1(x - \hat{x})^2] \exp \left(\frac{-j2\pi v_1 x}{v_a} \right) dx \end{aligned} \quad (A3)$$

and the expression for the received signal $r(t)$ from Eq (6) is

$$\begin{aligned}
 g(t, \hat{x}) = & \sum_n \text{rect} \left(\frac{t-nT}{\tau_0} \right) \int \text{rect} \left(\frac{x}{X} \right) \left\{ \sum_n A(nT) \text{rect} \left[\frac{t-2R_0/c-nT+x/v_a}{\tau} \right] \exp [jb(t-2R_0/c-nT+x/v_a)^2] \right. \\
 & \times \exp \left[-j \frac{2\pi v_0}{cR_0} (vnT-\eta_0)^2 \right] \exp [j2\pi v_0(t+x/v_a)] \exp [-j2\pi(v_0-v_1)(t+x/v_a)] + B \exp [j2\pi v_2(t+x/v_a)] \Big\} \\
 & \times \exp [jb_1(x-\hat{x})^2] \exp (-j2\pi v_1 x/v_a) dx .
 \end{aligned} \quad (A4)$$

The laser pulse width τ_0 is chosen sufficiently smaller than the bandwidth of the radar pulses ($\tau_0 < b\tau/\pi$) so that the temporal variation of the signal in the above equation can be neglected over the duration of each light pulse. With this approximation Eq. (A4) can be rewritten as follows:

$$\begin{aligned}
 g(t, \hat{x}) = & \sum_n \text{rect} \left(\frac{t-nT}{\tau_0} \right) \exp(j2\pi v_1 t) \int \text{rect} \left(\frac{x}{X} \right) \left\{ \sum_n A(nT) \text{rect} \left[\frac{x/v_a-2R_0/c}{\tau} \right] \exp [jb(x/v_a-2R_0/c)^2] \right. \\
 & \times \exp \left[\frac{-j2\pi v_0}{cR_0} (vnT-\eta_0)^2 \right] + B \exp [j2\pi(v_2-v_1)x/v_a] \exp [j2\pi(v_2-v_1)nT] \Big\} \exp [jb_1(x-\hat{x})^2] dx .
 \end{aligned} \quad (A5)$$

We can assume that $|X| > v_a \tau + 2R_0 v_a/c$, i.e., the range delay is such that the reflected radar pulse is entirely in the aperture of the AOD at the instant the laser is pulsed. This can be ensured by appropriately setting the relative phase of the radar and laser pulse repetition frequencies. The parameter b_1 (the focusing power of the optical system in the x direction) is set equal to b/v_a^2 . Finally, the offset frequency v_2-v_1 is made to be equal to m/T , where m is an integer. This can be accomplished by deriving v_1-v_2 from the pulse repetition frequency $1/T$. The function $g(\hat{x}, t)$ can then be written as follows:

$$\begin{aligned}
 g(\hat{x}, t) = & \sum_n \text{rect} \left(\frac{t-nT}{\tau_0} \right) \left\{ A(nT) \exp \left[-j \frac{2\pi v_0}{cR_0} (vnT-\eta_0)^2 \right] \exp \left(-j \frac{b}{v_a^2} \hat{x}^2 \right) \exp \left(j \frac{4bR_0^2}{c^2} \right) \int \text{rect} \left(\frac{x/v_a-2R_0/c}{\tau} \right) \right. \\
 & \times \exp \left[-j \frac{2b}{v_a^2} (\hat{x}-2R_0 v_a/c) x \right] dx + B \exp [j2\pi(v_2-v_1)\hat{x}/v_a] \exp [j\pi^2(v_2-v_1)^2/b] \Big\} = \sum_n \text{rect} \left(\frac{t-nT}{\tau_0} \right) \\
 & \times \left\{ A(nT) v_a \tau \text{sinc} \left[\frac{b\tau}{\pi v_a} (\hat{x}-2R_0 v_a/c) \right] \exp \left[-j \frac{2\pi v_0}{cR_0} (vnT-\eta_0)^2 \right] \exp \left[-j \frac{2b}{v_a^2} (\hat{x}-2R_0 v_a/c) (2R_0 v_a/c) \right] \right. \\
 & \times \exp \left[-j \frac{b}{v_a^2} (\hat{x}^2-4R_0^2 v_a^2/c^2) \right] + B \exp [j2\pi(v_1-v_2)\hat{x}/v_a] \exp [j\pi^2(v_2-v_1)^2/b] \Big\} .
 \end{aligned} \quad (A6)$$

We substitute the above expression in Eq. (14):

$$\begin{aligned}
 I(\hat{x}, \hat{y}, nT) = & T(\hat{x}, \hat{y}) \int_{t-nT}^{t+nT} \sum_n \text{rect} \left(\frac{t-nT}{\tau_0} \right) \left\{ |A(nT)|^2 v_a^2 \tau^2 \text{sinc}^2 \left[\frac{2b\tau}{v_a} (\hat{x}-2R_0 v_a/c) \right] + |B|^2 + 2B A(nT) v_a \right. \\
 & \times \text{sinc} \left[\frac{b\tau}{\pi v_a} (\hat{x}-2R_0 v_a/c) \right] \cos \left[\frac{2\pi(v_2-v_1)\hat{x}}{v_a} + \frac{2\pi v_0}{cR_0} (vnT-\eta_0)^2 + \phi \right] \Big\} dt = 2T(\hat{x}, \hat{y}) B v_a \tau \tau_0 A(nT) \\
 & \times \text{sinc} \left[\frac{b\tau}{\pi v_a} (\hat{x}-2R_0 v_a/c) \right] \cos [2\pi(v_2-v_1)\hat{x}/v_a + 2\pi v_0(vnT-\eta_0)^2/cR_0 + \phi] + \text{bias terms} .
 \end{aligned} \quad (14)$$

In the above equation, the phase terms from Eq. (A6) that do not depend on the index n have been lumped in the parameter ϕ .

8. APPENDIX B: DERIVATION OF EQUATION (15)

$$Q(\hat{x}, \hat{y}) = \sum_n Q(\hat{x}, \hat{y} - n\Delta y + N\Delta y, nT) \quad (15)$$

Substitute Eq. (14) in the above equation and the expression for the transmittance of the mask from Eq. (12):

$$\begin{aligned} Q(x, y) &= \sum_n 2Bv_\alpha \tau_{\tau_0} \left\{ \frac{1}{2} + \frac{1}{2} \right. \\ &\quad \times \cos \left[2\pi u_0 \hat{x} - \pi b_2 (\hat{y} - (n - N)\Delta y)^2 / \hat{x} \right] \\ &\quad \times A(nT) \operatorname{sinc} \left[\frac{b\tau}{\pi v_\alpha} (\hat{x} - 2R_0 v_\alpha / c) \right] \\ &\quad \times \cos \left[2\pi(v_2 - v_1)\hat{x} / v_\alpha + 2\pi v_0(vnT - \eta_0)^2 / cR_0 + \phi \right] \quad (B1) \end{aligned}$$

The spatial frequency u_0 on the mask must be selected so that the sum frequency $u_0 + (v_2 - v_1)/v_\alpha$ of the two cosine functions in the above equation is sufficiently large so that the signal term that is riding on the sum frequency can be separated from the bias terms in the spatial frequency domain. To accomplish this, we set $u_0 = (v_2 - v_1)/v_\alpha$. We write the product of the cosines as a sum, and the term of interest in Eq. (B1) takes the following form:

$$\begin{aligned} Q(\hat{x}, \hat{y}) &= (Bv_\alpha \tau_{\tau_0} / 2) \operatorname{sinc} \left[(b\tau / \pi v_\alpha) (\hat{x} - 2R_0 v_\alpha / c) \right] \\ &\quad \times \sum_n A(nT) \cos \left[4\pi u_0 \hat{x} - \pi b_2 \hat{y}^2 / \hat{x} \right. \\ &\quad \left. + 2\pi b_2 (n - N)\Delta y \hat{y} / \hat{x} - \pi b_2 \Delta y^2 (n - N)^2 / \hat{x} + 2\pi v_0 v^2 T^2 n^2 / cR_0 \right. \\ &\quad \left. - 4\pi v_0 n T \eta_0 / cR_0 + 2\pi v_0 \eta_0^2 / cR_0 + \phi \right] \quad (B2) \end{aligned}$$

The focusing of the SAR image in the azimuth direction (η) is accomplished by the summation over n in the above equation. The summation will result in a well-focused image in the \hat{y} direction if the terms containing n^2 in the argument of the cosine are canceled. To accomplish this, the constant b_2 of the mask is set appropriately.

$$\begin{aligned} (2v_0 v^2 T^2 / cR_0) n^2 &= (b_2 \Delta y^2 / \hat{x}) n^2 \rightarrow \\ \rightarrow b_2 &= \frac{2v_0 v^2 T^2 \hat{x}}{cR_0} = \frac{2v_0 v^2 T^2}{cR_0} (2R_0 v_\alpha / c) \\ &= \frac{4v_0 v^2 T^2 v_\alpha}{c^2 \Delta y^2} \quad (B3) \end{aligned}$$

Equation (B3) and the value we previously chose for u_0 provide all the information necessary to construct the mask that is used in the system. Equation (B3) is substituted in Eq. (B2), and all inconsequential phase terms are included in the parameter ϕ . Equation (B2) then reduces to the following equation:

$$\begin{aligned} Q(\hat{x}, \hat{y}) &= (Bv_\alpha \tau_{\tau_0} / 2) \operatorname{sinc} \left[(b\tau / \pi v_\alpha) (\hat{x} - 2R_0 v_\alpha / c) \right] \\ &\quad \times \sum_n A(nT) \cos \left[4\pi u_0 x + (4\pi v_0 v^2 T^2 / c\Delta y R_0) \right. \\ &\quad \left. \times (\hat{y} - (\eta_0 \Delta y / vT) + N\Delta y) n + \phi \right] \quad (B4) \end{aligned}$$

The summation over n in Eq. (B4) is recognized as the discrete cosine transform¹³ of the antenna pattern $A(nT)$. In order to get an approximate result, we can assume that the antenna pattern is so broad in the azimuth direction that its main lobe uniformly illuminates a distance vNT on the ground. Therefore, $A(nT)$ is approximately a constant for $0 > n > N$, and Eq. (B4) can be evaluated as follows¹³:

$$\begin{aligned} Q(\hat{x}, \hat{y}) &= (Bv_\alpha \tau_{\tau_0} / 2) \operatorname{sinc} \left[(b\tau / \pi v_\alpha) (\hat{x} - 2R_0 v_\alpha / c) \right] \\ &\quad \times \frac{\sin \left[(2N\pi v_0 v^2 T^2 / c\Delta y R_0) (\hat{y} - (\eta_0 \Delta y / vT) + N\Delta y) \right]}{\sin \left[(2\pi v_0 v^2 T^2 / c\Delta y R_0) (\hat{y} - (\eta_0 \Delta y / vT) + N\Delta y) \right]} \\ &\quad \times \cos (4\pi u_0 x + \phi) \quad (B5) \end{aligned}$$

The ratio of the sines in Eq. (B5) can be approximated by a sinc function for large N , and this leads us to the final form of Eq. (15):

$$\begin{aligned} Q(\hat{x}, \hat{y}) &= (NBv_\alpha \tau_{\tau_0} / 2) \operatorname{sinc} \left[(b\tau / \pi v_\alpha) (\hat{x} - 2R_0 v_\alpha / c) \right] \\ &\quad \times \operatorname{sinc} \left[(2Nv_0 v^2 T^2 / c\Delta y R_0) (\hat{y} - (\eta_0 \Delta y / vT) + N\Delta y) \right] \\ &\quad \times \cos (4\pi u_0 x + \phi) \quad (15) \end{aligned}$$

9. REFERENCES

1. D. A. Ausherman, *Opt. Eng.* 19(2), 157(1980).
2. T. M. Turpin, *Proc. IEEE* 69(1), 79(1981).
3. W. T. Rhodes, *Proc. IEEE* 69(1), 65(1981).
4. E. N. Leith, *Proc. IEEE* 59(9), 1305(1971).
5. I. Cindrich, J. Marks, and A. Klooster, *Proc. SPIE* 128, 128(1977).
6. A. Korpel, *Proc. IEEE* 69(1), 48(1981).
7. K. Bromley, M. Monahan et al., *Proc. SPIE* 118, 118(1977).
8. M. Blouke et al., *Proc. SPIE* 290, 6(1981).
9. D. Casasent, *Proc. IEEE* 65, 143 (Jan. 1977).
10. J. Hall, in *Applied Optics and Optical Engineering*, Vol. VIII, Chap. 8, Academic Press, New York (1980).
11. D. Psaltis and K. Wagner, *Proc. SPIE* 271, 51(1981).
12. A. W. Rihaczek, *Principles of High Resolution Radar*, McGraw Hill, New York (1969).
13. A. Oppenheim and R. Schaffer, *Digital Signal Processing*, Prentice Hall, Englewood Cliffs, New Jersey (1975).

PROGRAMMABLE REAL-TIME ACOUSTO-OPTIC/CCD SAR PROCESSOR

Michael Haney*, Kelvin Wagner, Demetri Psaltis

California Institute of Technology
Department of Electrical Engineering
Pasadena, California 91125**Abstract**

The theory of operation of the Real-Time Acousto-Optic SAR Processor is reviewed and recent experimental results are presented. The results include a demonstration of the real-time imaging capability of the processor with simulated radar signals. An advanced version of this processor is then described in which a programmable reference function is entered via a second acousto-optic device to eliminate the need for a 2-D SLM. In this implementation the reference function is updated by electronic means to give the processor the flexibility to adapt rapidly to changes in the parameters of the radar/target geometry.

Introduction

An acousto-optic processor which employs time and space integrating techniques to form Synthetic Aperture Radar (SAR) images in real time has been described previously^{1,2}. In this architecture the unfocussed radar signals are entered into the optical system via an acousto-optic device (AOD) and the focussed image is formed on a 2-D CCD. The range and azimuth focussing are performed with spatial and temporal integrations, respectively. This processor has been evaluated with input signals that simulate the radar returns that are received from individual point scatterers in the target field. Since the system is linear, the processing of signals from a point scatterer constitutes a measurement of the impulse response, which completely characterizes the focussing ability of the processor. The results of this evaluation are presented in the next section following a brief review of the theory of operation of the processor.

In some applications of SAR the real-time processor must adapt rapidly to changes that occur in the radar/target geometry in order to continue producing a well focussed image. One way to accomplish this in the AO/CCD architecture is through the use of a real time 2-D SLM in place of the fixed mask on which the reference chirp signals are recorded. In this paper we describe an alternate and preferable approach in which an interferometric scheme is employed to produce a programmable reference function and to eliminate the need for a 2-D SLM. In this advanced version of the processor we exploit the fact that the desired 2-D reference function is a 1-D linear FM signal whose scale varies in the second dimension to account for the range/azimuth coupling of the SAR geometry; thus a 1-D SLM can be used to enter the reference function into the processor while the scaling is accomplished by an appropriate lens configuration. In this paper the programmable architecture is described and the results of the experimental demonstration of the interferometric generation of reference functions are presented.

Time and Space Integrating SAR Architecture

A typical side-looking radar/target geometry is depicted in Figure 1. The radar is carried on an aircraft or satellite flying at a constant velocity parallel to the ground. The radar illuminates an area of the ground to the side of the flight path so that a swath is swept out as shown in Figure 1. The illuminating radiation is temporally modulated by periodic LFM pulses so that, as the beam passes over a point scatterer, a sequence of LFM radar returns will be received from it. The changing radar/target aspect angle leads to a slow phase modulation over a sequence of pulses received from a point scatterer that is superimposed on the high-speed quadratic phase modulation within each pulse. This slower phase modulation is approximately quadratic with time and has a chirp rate that is inversely proportional to the range of the point scatterer. The chirp rate is sufficiently low that the phase change within a single pulse can be considered negligible. Consequently, for a single point scatterer located at (x_0, y_0) , the 1-D sequence of radar returns can be expressed in a 2-D format³ as:

$$r(x, y) = A[\exp[j\{k_1(x-x_0)^2/y_0 + k_2(y-y_0)^2\}]] \quad (1)$$

where A , k_1 , and k_2 are constants and the variables x and y correspond to the range and azimuth coordinates on the ground. To produce a focussed image $g(x, y)$ of the point

* M. Haney is with General Dynamics Corp., Pomona, Ca., and is currently attending Caltech on a corporate sponsored doctoral fellowship.

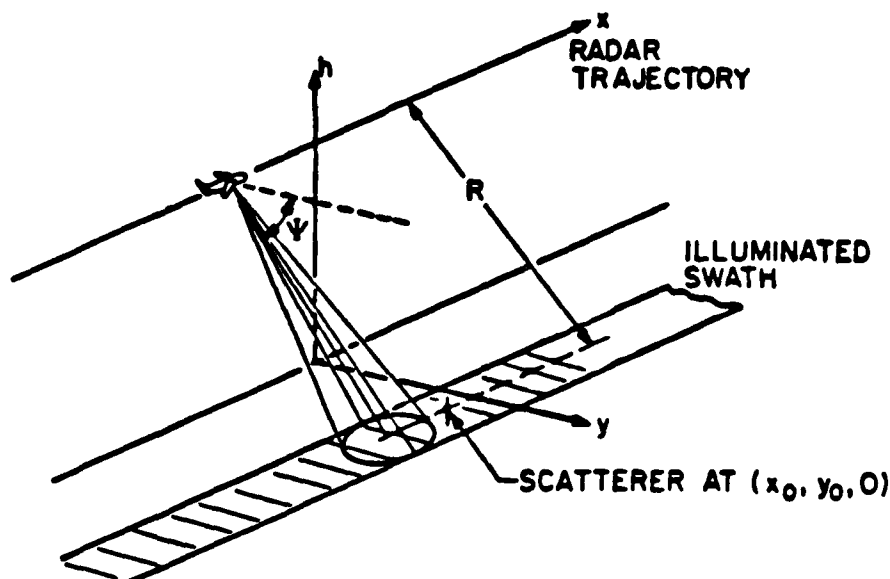


Figure 1. Typical SAR Geometry

scatterer, the signal $r(x,y)$ must be correlated with a kernel $h(x,y)$, equal to $r(x,y)$, but shifted to the origin. Thus we have:

$$g(x,y) = \iint r(u,v) h^*(u+x, y+v) du dv \quad (2)$$

where:

$$h(x,y) = \exp[j(k_x x^2 / y_0 + k_y y^2)] \quad (3)$$

The focussing of SAR images is, as evidenced by Equation (2), a two dimensional processing problem. Optical Signal Processing techniques, being inherently 2-D, were therefore recognized, at the introduction of SAR, to be applicable. In the non-real time, film-based optical processing approach the radar return signal $r(x,y)$ is recorded on film and the image is generated by running the developed film through a 2-D correlator which then performs a spatial integration in both the range and azimuth dimensions simultaneously.

Equation (2) can be rewritten in the following manner:

$$g(x,y) = \int \{ \int r(u,v) \exp[j(k_y(v+y)^2)] dv \} \exp[j(k_x(u+x)^2/y_0)] du \quad (4)$$

Equation (4) reveals that, the two integrations do not necessarily have to be performed simultaneously, but instead can be performed as a cascade of 2 one dimensional integrations. This feature permits the AO/CCD processor to use a 1-D SLM as an input device and process the reflection of each radar pulse one at a time. Real time operation is achieved by performing a spatial integration in the range dimension on each of the radar returns as they are received, and a temporal integration in the azimuth dimension as the range focussed radar returns accumulate on the detector at the output of the processor.

The range processing technique is depicted in Figure 2. The radar returns are mixed to the center frequency of an acousto-optic device (AOD) and applied to its input. The AOD is illuminated with a beam from a laser diode that is collimated in the x direction. The laser diode is pulsed in synchronism with the radar PRF, and has a pulsewidth which is less than the inverse of twice the bandwidth of the radar chirp to essentially freeze the moving diffraction grating in the AOD. The diffraction grating induced by each point scatterer is a one dimensional Fresnel zone plate which focuses the light onto a CCD detector to effect the pulse compression in the range dimension.

The azimuth pulse compression is performed by correlation of the quadratic phase history of the radar returns from each point scatterer with a mask on which the matching phase history is recorded for each range bin. This technique is depicted in Figure 3. The phase of the radar signals is detected by interfering the range focussed light with a plane wave generated by applying a sinusoidal reference signal to the AOD in addition to

the radar signal. The light from the AOD is collimated in the y direction to uniformly illuminate the mask which is placed directly in front of the 2-D CCD detector. The CCD is operated in the shift-and-add mode, in the y direction, at a shift rate equal to the PRF of the radar, thus giving the effect of a moving detector behind the stationary mask. As the photogenerated charge on the detector moves completely across the mask, the charge that builds up on the detector corresponds to the completed azimuth correlation. The image, which has now been focussed in both range and azimuth, is read out by the CCD at a line rate equal to the PRF of the radar. The complete AO/CCD Real-time SAR Processor is shown schematically in Figure 4 and an example of a reference mask is given in Figure 5.

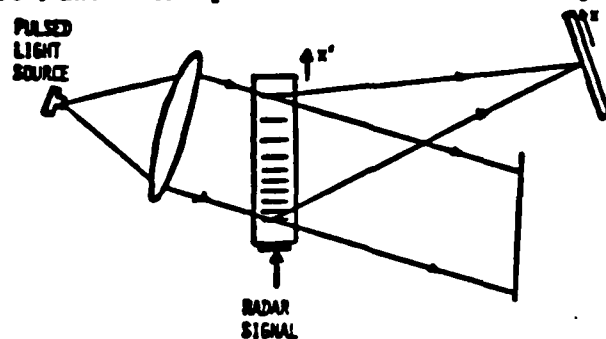


Figure 2. Acousto-optic processor that performs the focussing of the SAR image in range.

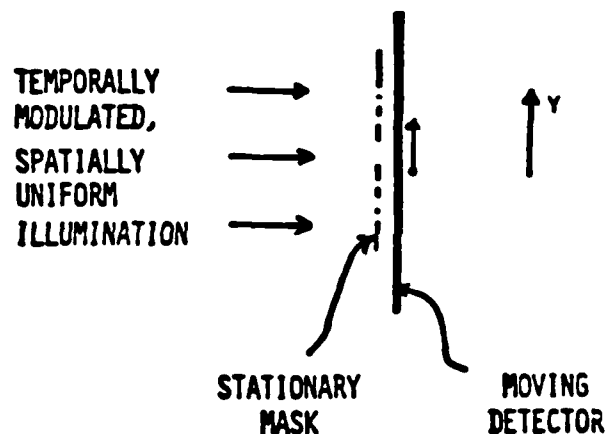


Figure 3. Time integrating optical processor that performs the focussing of the SAR image in azimuth.

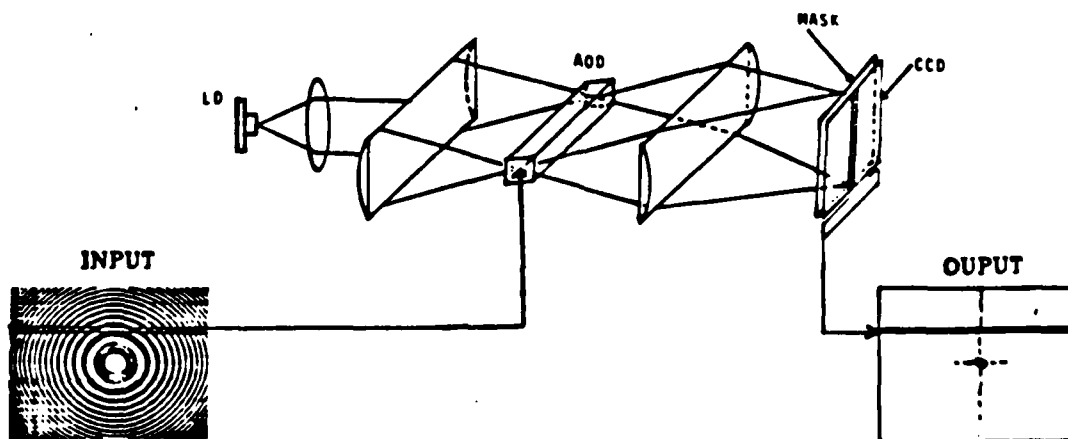


Figure 4. Real-time Acousto-optic/CCD SAR Processor

The real-time imaging ability of the AO/CCD system has been demonstrated by measuring the impulse response of the processor. A block diagram of the electronics used in these measurements is shown in Figure 6 interfaced to the optical processor. The SAR Impulse Response Generator simulates the radar signal that would be received from a single point scatterer in the target field. A SAW device is used to produce a sequence of LFM pulses at a fixed PRF. The quadratic phase history of a sequence of radar returns from a single point scatterer in a particular geometry is calculated and stored in digital memory. The pulses from the SAW device pass through a digitally controlled phase shifter which imparts the stored phase history to the sequence. The simulated radar returns are then summed to the reference signal and applied to the AOD of the processor. The electronic signals that trigger the LD and CCD, and produce the reference beam for the azimuth correlation are all generated in synchronism with the Impulse Response Generator. The isometric display block grabs a frame of the scrolling CCD output and displays the focussed image of the point scatterer in either a 2-D or 3-D perspective format for analysis.

Figure 6. Impulse Response Experiment

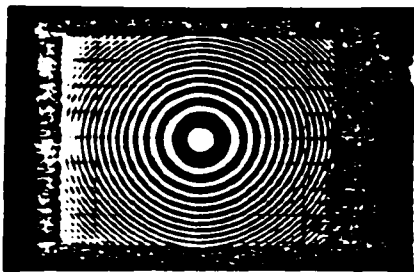


Figure 7. Impulse Response Generator output.

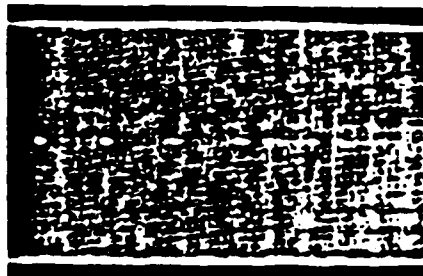


Figure 8. AO/CCD Processor output: Real-time imaging of simulated radar point scatterers.

Interferometrically Generated Azimuth Reference Function

In some applications the parameters of the radar/target geometry, such as the velocity and altitude of the aircraft, and direction in which the antenna is pointed, change dynamically. The real-time SAR imager must have the ability to adapt rapidly to such changes in order to continuously provide a well focussed image. Since the impulse response of the AO/CCD processor in the azimuth direction is determined by the mask on which the reference chirps are recorded, the mask could be replaced by a 2-D SLM on which the proper reference function could be written and updated as needed. However, examination of the required 2-D reference function (see Figure 5) reveals that, when the overall range to the target is large when compared to the swath width, then the reference function is approximated well by a 1-D baseband LFM signal in the azimuth direction, whose scale varies in the range direction to account for the range/azimuth coupling of the geometry. This suggests that a 1-D SLM can be used to input the reference function to the processor and the range/azimuth coupling can be introduced by an appropriate anamorphic lens arrangement.

An interferometric scheme in which the 2-D azimuth reference function is generated using a 1-D input SLM is shown schematically in Figure 9. An electronic chirp signal is applied to an AOD along with a sinusoidal reference signal. After the acoustic signals have

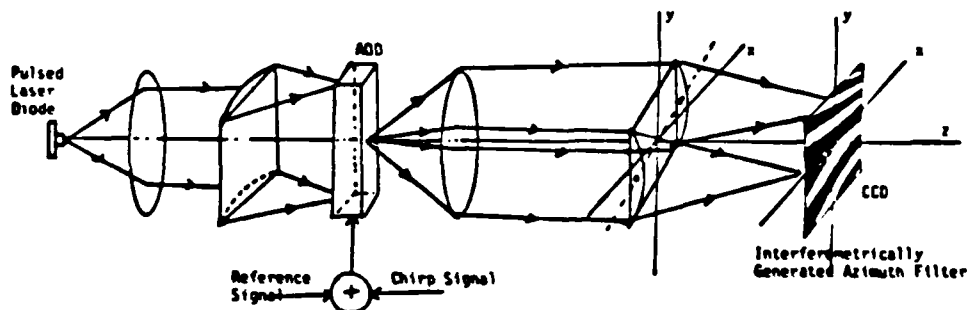


Figure 9. Optical Set-up for the Interferometric Generation of the SAR azimuth filter

propagated to the center of the AOD, the laser diode is pulsed to freeze the moving diffraction gratings in the same manner as the radar signal is sampled in the AOD of the SAR processor described previously. An anamorphic lens arrangement is used to spread the diffracted light uniformly in the x direction at the detector plane. The undiffracted light (not shown in Figure 9) is blocked and not used. In the y direction the diffracted light from the sinusoidal reference signal in the AOD will be a plane wave that is transformed into a cylindrical wave at the detector; the light diffracted by the chirp signal will be a cylindrical wave that is transformed into a cylindrical wave of different curvature at the detector. The waves from the chirp and reference signals will have different radii of curvature and therefore the resulting interference pattern on the detector will be a LFM signal in the y direction. The scaling that is required in the x direction to account for the range/azimuth coupling is accomplished by tilting a cylindrical lens about the y axis, as shown in Figure 9.

Figure 10 is a photograph of an interferometrically generated azimuth filter produced with the system described above. The 2-D CCD camera was operated in standard video mode and its output displayed on a monitor for data acquisition. During the integration period of each video frame the AOD was loaded and illuminated several hundred times to produce a high contrast ratio in the reference function. A comparison of Figures 10 and 5 shows that, aside from differences in levels of exposure, the interferometrically generated

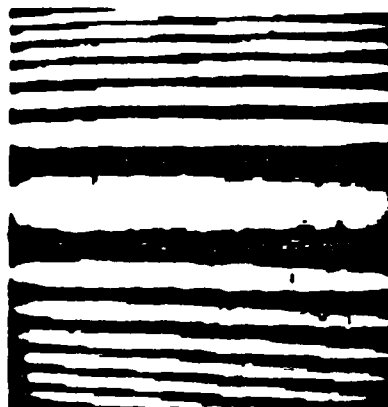


Figure 10. Interferometrically generated azimuth reference function.

function has nearly the same azimuth bandwidth and range/azimuth coupling parameter as the fixed mask.

Programmable Architecture

The interferometric technique for SAR azimuth filtering can be integrated with the range processing technique of Figure 2 to fashion a maskless programmable real-time SAR processing architecture. One approach is a two arm interferometer in which the light that is modulated by the range focussing AOD follows a different path to the output CCD detector than the light that forms the azimuth filter. Alternatively, it is possible to cascade the two systems so that the interfering light beams follow the same path through the optical system, thereby minimizing the stability requirement. The cascaded system is shown in Figure 11. As in the original processor, the radar signal is range focussed and

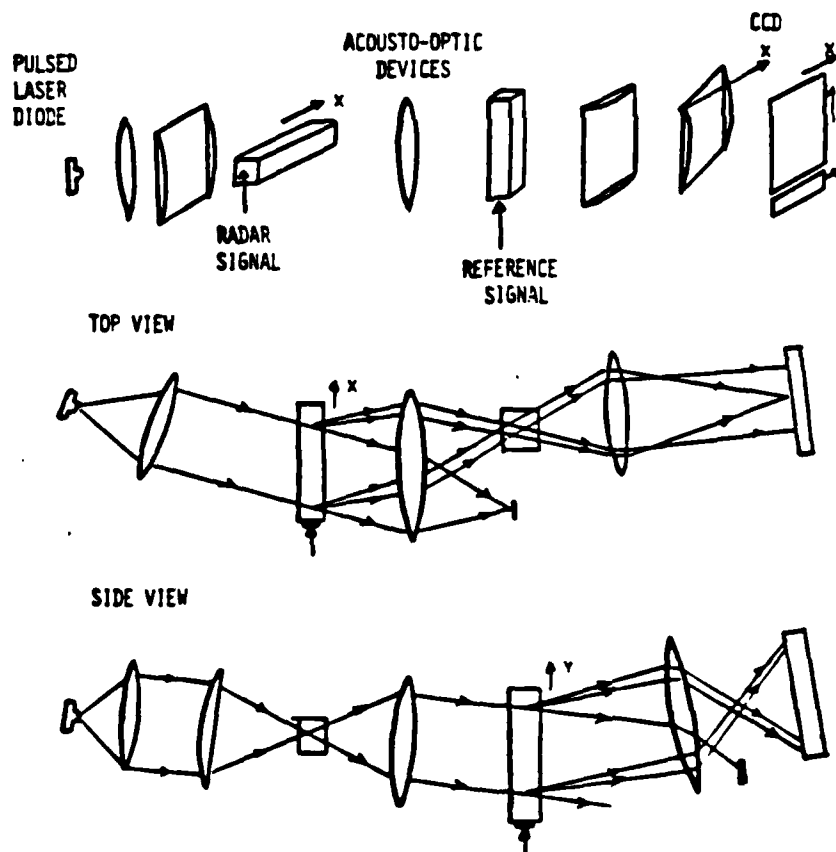


Figure 11. Programmable Architecture

its phase is detected interferometrically at the detector. Since the light is now also diffracted by the second AOD, containing the azimuth reference signal, it is effectively multiplied by the interferometrically generated reference function. Operation of the CCD in the shift-and-add mode will complete the necessary azimuth correlation as in the non-programmable architecture.

The programmable architecture gives the AO/CCD processor flexibility to adapt to changes in the radar/target geometry by electronically changing the starting frequency and chirp rate of the azimuth reference signal. Equally significant is the fact that this architecture allows the azimuth correlation to be performed by shifting the reference function incrementally in the AOD rather than shifting and adding the charge on the CCD. This capability is useful for performing multiple looks, i.e. forming several images of the same area independently and adding them to reduce speckle. Furthermore, shifting the reference signal in the AOD is useful for Spot-light Mode SAR, in which it is desired to form the image in a frame-by-frame rather than in a scrolling manner.

Acknowledgements

The work reported in this paper is funded by NASA, General Dynamics, and the Air Force Office of Scientific Research. The authors wish to thank T. Bicknell, J. Yu, and W. Farr for their assistance in the experiments.

References

1. D. Psaltis and K. Wagner, "Real-time Optical SAR Processor," Opt. Eng., Vol. 21, p. 822, 1982.
2. D. Psaltis, M. Haney, and K. Wagner, "Real-Time Synthetic Aperture Radar Processing," Proc. NASA Conf. on Optical Information Processing, (Hampton, Va., 1983).
3. L. J. Cutrona, et al, "On The Application of Coherent Optical Processing Techniques to Synthetic Aperture Radar," Proc. IEEE, 54, 1826 (1966).

ACOUSTO-OPTIC TECHNIQUES FOR REAL TIME SAR IMAGING

Michael Baney* and Demetri Psaltis

California Institute of Technology
Department of Electrical Engineering
Pasadena, California 91125

Abstract

Recent advancements in the development of the Real Time Acousto-optic SAR Processor are presented. In particular, the technique for introducing the azimuth reference function into the processor via an-acousto-optic Bragg cell is discussed. This approach permits the reference function to be stored in electronic memory, thus giving the processor the flexibility needed to adapt rapidly to changes in the radar/target geometry. The architecture is described and results are presented which show the applicability of the technique to both spot-light and strip-map SAR.

Introduction

The formation of high-resolution images using the Synthetic Aperture Radar (SAR) technique is one of the most successful applications of optical signal processing technology. Film-based coherent processors account for a major portion of the SAR images produced to date.^{1,2} Although film provides a convenient high-density data storage medium and can also serve as input spatial light modulator (SLM) for the optical processor, the need for chemical development of the exposed film precludes its use in real-time applications. Recently, advances in the technology of light sources, modulators, and detectors have led to optical SAR imagers which do not require film, can handle the computational load at real-time rates, and are relatively compact and low in power consumption.

An acousto-optic time-and-space integrating architecture for real-time SAR imaging has been described previously.^{3,4} In this architecture an acousto-optic device (AOD) is used to enter the unfocused radar signals into the optical system. The image is formed on a 2-D charge-coupled device (CCD) detector array. The range and azimuth focusing are performed with spatial and temporal integrations of light, respectively.

An advanced version of this processor is now being developed in which the azimuth reference function is entered into the processor via a second AOD⁵, instead of with a 2-D intensity mask or SLM as was done in the earlier architectures. With this technique the desired azimuth filter function can be synthesized electronically and updated at high rates.

In the next section of this paper we briefly review the signal processing requirements of a SAR imager. This discussion serves to illuminate the aspects of the real-time SAR problem that are exploited in the AO/CCD processor. The techniques used to accomplish the range and azimuth compressions are then described. The theory of operation of the complete architecture, in which the range and azimuth processing are combined interferometrically, is presented. Finally, we show the results of an impulse response characterization of the processor in which both framing and scrolling modes are demonstrated.

Signal Processing Requirements

In a SAR geometry, the result of the radar illumination and subsequent collection of the backscattered radiation is a mapping of the 2-D reflectivity of the target scene into its unfocused image. The unfocused image is a summation of windowed, weighted, asymmetrical 2-D quadratic phase functions, whose positions and weights are determined by the location and strength of the associated point scatterers of the target scene. The degree of asymmetry in the unfocused image of each point scatterer is determined by the parameters of the radar and the radar/target geometry. The 2-D filtering required to form the image is a linear operation. Therefore, in describing the processing, we need consider only a single point scatterer as the input. For a single point scatterer located at coordinates (x, y) ,

* M. Baney is with General Dynamics Corp., Pomona, Ca., and is currently attending Caltech on a corporate sponsored doctoral fellowship.

the 1-D sequence of radar returns is expressed in a 2-D format² as:

$$r(x,y) = A[\exp\{j[k_x(x-x_0)^2 + k_y(y-y_0)^2(1+ax_0)/R_0]\}] \quad (1)$$

where A , a , k_x , and k_y are constants determined by the parameters of the radar and the radar/target geometry. The variables x and y correspond to the range and azimuth coordinates on the ground, respectively. The origin of the ground coordinates is taken to be the center of the swath located at range R_0 , with $y=0$ defined to be the position directly abeam of the radar at the start of data collection. In Eq. (1) we assume $x_0 \ll R_0$.

In the conventional non-real-time optical approach, the function in Eq. (1) is recorded on photographic film as a Fresnel zone plate. The image is then formed by a coherent optical processor consisting of an anamorphic lens arrangement to correct for the range/azimuth coupling and the zone plate asymmetry, which focuses the collection of zone plates into a high resolution image. For each point scatterer, the required operation is a correlation with a kernel $h(x,y)$, equal to $r(x,y)$, but shifted to the origin. Thus we have:

$$g(x,y) = \iint r(u,v)h^*(u+x,y+v)dudv \quad (2)$$

where:

$$h(x,y) = \exp\{j[k_x x^2 + k_y y^2(1+ax_0)/R_0]\} \quad (3)$$

The required processing steps of a SAR imager are shown pictorially in Fig. 1.

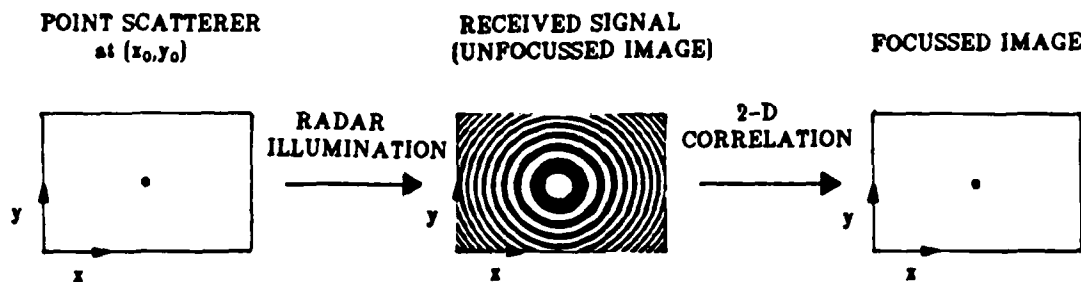


Figure 1. SAR Processing

In the film-based optical processing approach the image is generated by running the data film through a 2-D correlator which performs a spatial integration of light in both the range and azimuth dimensions simultaneously. However, the required integrations in the two dimensions do not necessarily have to be performed at the same time. This is shown by rewriting Eq. (2) in the following manner:

$$g(x,y) = \int \left[\int r(u,v) \exp(-j(k_x(u+x)^2) du \right] \exp(-j(k_y(1+ax_0)(v+y)^2/R_0) dv \quad (4)$$

Eq. (4) reveals that the two integrations can be performed as a cascade of two one dimensional integrations. This feature permits the AO/CCD processor to use a 1-D SLM as an input device and process the reflection of each radar pulse one at a time. Real-time operation is achieved by performing a spatial integration in the range dimension on each of the radar returns as they are received, and a temporal integration in the azimuth dimension as the range focused radar returns accumulate on the detector at the output of the processor.

Architecture Description

The range processing technique is depicted in Fig. 2. The radar returns are mixed to the center frequency of an AOD and applied to its input. The AOD is illuminated with a

beam from a laser diode that is collimated in the x direction. The laser diode is pulsed in synchronism with the radar PRF, and has a pulsewidth which is less than one half the inverse of the bandwidth of the radar chirp in order to freeze the moving diffraction grating in the AOD. The diffraction grating induced by each point scatterer is a one dimensional Fresnel zone plate which focuses the light onto a CCD detector to effect the pulse compression in the range dimension.

The interferometric technique by which the phase of each radar return is detected for the azimuth processor is also depicted in Fig. 2. A sinusoidal reference signal, which is coherent with the radar waveform, is added to the radar return signal and the sum is applied to the input transducer of the AOD. The light diffracted by the AOD due to the reference is a plane wave in the x direction which interferes with the range focused light at the detector plane to record its amplitude and phase. This technique has a significant advantage over other interferometric approaches in that it is insensitive to mechanical vibrations, since the interfering beams pass through the same optical elements. The azimuth pulse compression is performed by correlating the quadratic phase history of the radar returns from each point scatterer with a spatial intensity modulation which matches the phase history for each range bin. In the original architecture the spatial modulation was performed by a fixed intensity mask. The time integration technique for the azimuth compression is depicted in Fig. 3. The phase of the range focused radar signals is interferometrically detected as described above, resulting in a sampled temporal intensity modulation of the range focused light beam, occurring over the sequence of radar returns. The light from the AOD is collimated in the y direction to uniformly illuminate the mask whose transmittance is equal to the azimuth phase history and is placed directly in front of, or imaged onto the 2-D CCD detector. The CCD is operated in the shift-and-add mode, in the y direction, at a shift rate equal to the PRF of the radar, thus giving the effect of a moving integrating detector behind the stationary mask. As the photogenerated charge on the detector moves across the mask, the charge that builds up on the detector corresponds to the azimuth correlation. The image, which has now been focused in both range and azimuth, is read out by the CCD at a line rate equal to the PRF of the radar. Electronic post-processing is used to remove bias terms that form due to the interferometric detection. The complete AO/CCD Real-time SAR Processor is shown schematically in Fig. 4 and an example of a reference function, in the form of an intensity mask, is given in Fig. 5.

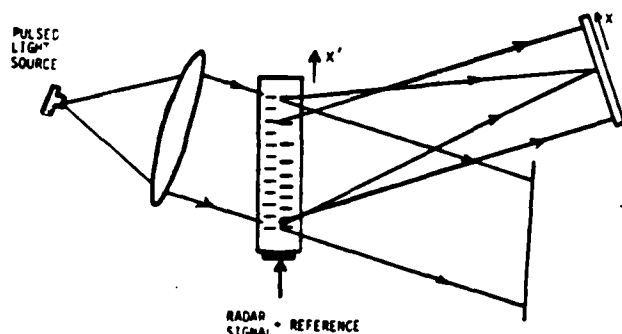


Figure 2. Acousto-optic processor that performs the range focusing of the SAR image and phase detection for the azimuth correlator.

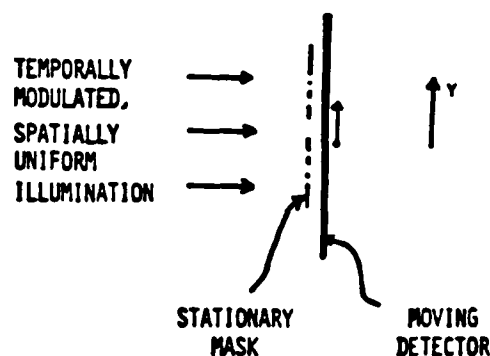


Figure 3. Time integrating optical processor that performs the focusing of the SAR image in azimuth.

In some applications, the parameters of the radar/target geometry, such as the velocity and altitude of the aircraft, and direction in which the antenna is pointed, may change dynamically. The real-time SAR imager must have the ability to adapt rapidly to such changes in order to continuously provide a well focused image. To accomplish this with the system of Fig. 4, the mask could be replaced by a real-time 2-D SLM on which the proper reference function is written and updated as needed. Examination of the required 2-D reference function (see Fig. 5) reveals that, when the overall range to the target is large when compared to the swath width, the reference function is approximated well by a 1-D baseband LFM signal in the azimuth direction, whose scale varies linearly in the range direction to account for the range/azimuth coupling of the geometry. Therefore, a 1-D SLM

can be used to input the reference function to the processor and the range/azimuth coupling can be introduced by an appropriate anamorphic lens arrangement to provide the scaling in the orthogonal direction. Fig. 6 is an example of an interferometrically

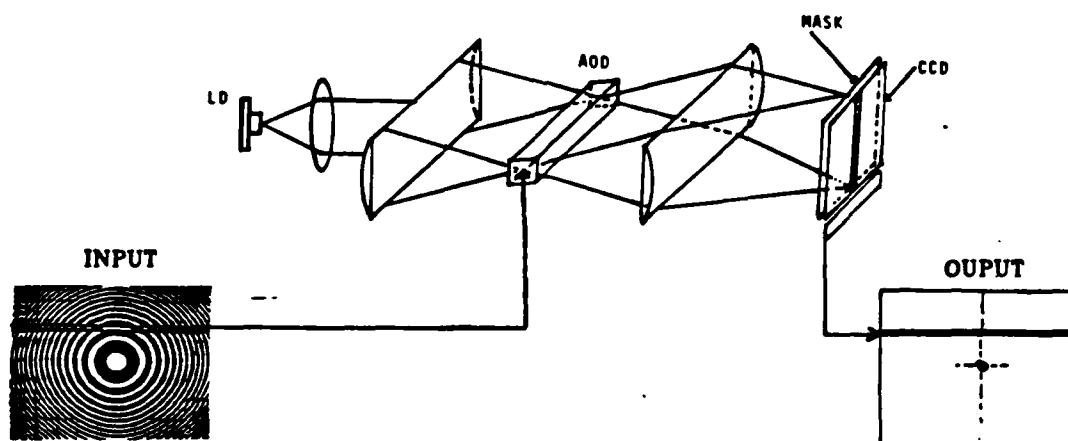


Figure 4. Real-time Acousto-optic/CCD SAR Processor.

generated azimuth filter whose characteristics were chosen to match those of the intensity mask of Fig. 5. The optical set-up used to produce the data of Fig. 6 is shown schematically in Fig. 7. An electronic chirp signal is applied to an AOD along with a sinusoidal reference signal. After the acoustic signals have propagated to the center of the AOD, the laser diode is pulsed to freeze the moving diffraction gratings in the same manner as the radar signal is sampled in the AOD of the SAR processor described previously. An anamorphic lens arrangement is used to spread the diffracted light uniformly in the x direction at the detector plane. The undiffracted light (not shown in Fig. 7) is blocked and not used. In the y direction the diffracted light from the sinusoidal reference signal in the AOD is a plane wave that is transformed into a cylindrical wave at the detector plane. The light diffracted by the chirp signal is a cylindrical wave that is transformed into a cylindrical wave of different curvature at the detector plane. The waves from the chirp and reference signals have different radii of curvature and therefore the resulting interference pattern on the detector will be approximately equal to a LFM signal in the y direction. The scaling that is required in the x direction to account for the range/azimuth coupling is accomplished by tilting a cylindrical lens about the y axis, as shown in Fig. 7. A comparison of Figs. 6 and 5 shows that, aside from differences in levels of exposure, the interferometrically generated function has nearly the same azimuth bandwidth and range/azimuth coupling parameter as the fixed mask. An architecture using the interferometrically generated reference filter has the flexibility to adapt to changes in the radar/target geometry by changing the characteristics of the electronically generated chirp. The parameters that may be changed include the starting frequency, the starting time, and the chirp rate of the azimuth reference signal.

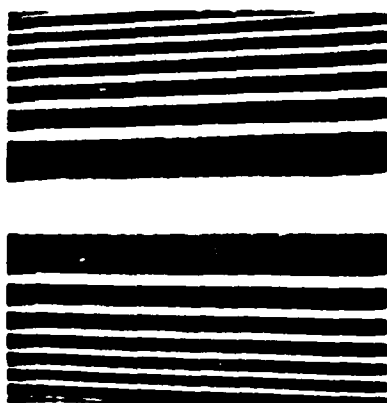


Figure 5. Azimuth reference mask.



Figure 6. Interferometrically generated azimuth reference function.

A programmable real-time SAR processing architecture is synthesized by integrating the interferometric technique for SAR azimuth filtering with the range processing technique of Fig. 2. One possible approach is an additive interferometric architecture in which the AOD containing the radar waveform is placed in one arm of a Mach-Zehnder interferometer and the AOD containing the azimuth reference is placed in the second arm and oriented orthogonally to the first AOD. This approach, however, is subject to the same sensitivity to mechanical vibration that plagues all multiple path interferometers.

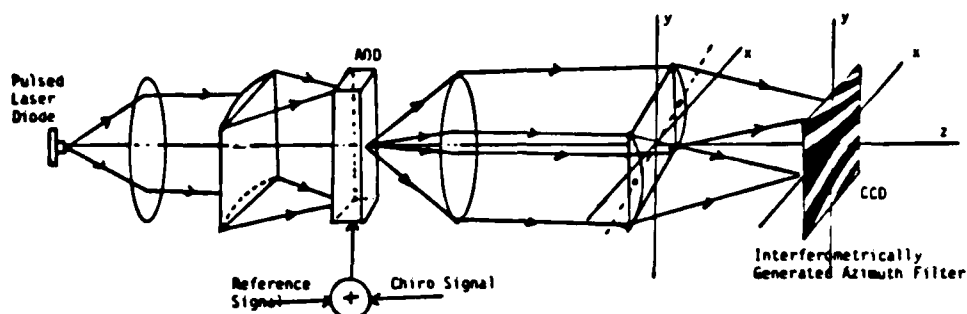


Figure 7. Optical Set-up for the Interferometric Generation of the SAR azimuth filter

An alternate and preferable approach is a multiplicative interferometric architecture in which the radar and azimuth filter systems are cascaded. In this implementation the interfering beams all pass through the same optics and thus the stability problems are minimized. The cascaded system is shown schematically in Fig. 8. As in the original

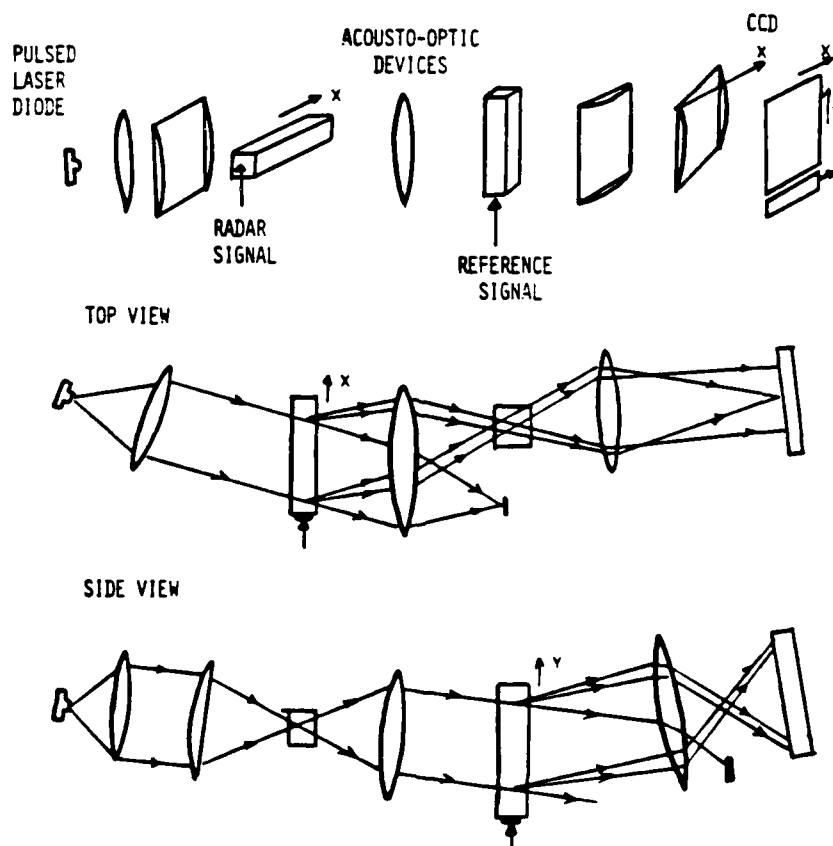


Figure 8. Programmable Architecture

processor, the radar signal is range focused and its phase is detected interferometrically at the detector by a reference beam that is generated by the same AOD. However, since the diffracted light from the first AOD is also diffracted by the second AOD, which contains the azimuth reference signal plus a sinusoidal signal, it is multiplied by the interferometrically generated reference function. This implementation thus achieves the same multiplication needed to perform the azimuth compression as in the original non-programmable architecture.

This architecture offers two possible modes of operation. Operation of the CCD in the shift-and-add mode, while repeating the same azimuth filter signal for each radar return, will complete the necessary azimuth correlation as in the original processor. This mode is used for side-looking SAR in which the output image is a long strip. Alternatively, the azimuth correlation may be performed by electronically shifting the reference function incrementally in the AOD and using a nonshifting integrating CCD detector array. In this mode the image is generated in a framed format that is most useful for spotlight mode SAR.

The operation of the multiplicative architecture is best understood by considering the two AODs as stationary 1-D amplitude/phase gratings. This is justified because the pulse-width of the laser diode is short enough to sample the moving diffraction patterns in the AODs. Consider as input to the processor the radar returns from a point scatterer at ground coordinates (\bar{x}, \bar{y}) with complex strength A . The corresponding range and cross-range (azimuth) coordinates in the Processor are x , and y , respectively. For simplicity we assume the x and y coordinates in the output plane of the processor to be the same as in the corresponding AODs, i.e., the optical magnification is 1. The strength of the reference signal in the range processing AOD is B , which is assumed to be real without any loss of generality. The radar returns are mixed down to the center frequency of the AOD, ω , which is also the frequency of the reference signal. A total of N returns are received from the point scatterer. With these assumptions, the 1-D diffraction pattern induced along the x axis in the first AOD, by the n th radar return, is expressed in the following way:

$$P_1(x, y, n) = [A \exp(j(k_1(x - x_0) + \bar{k}_1(1 + \alpha x_0)(n - y)/\Delta y) + R_0) \text{rect}((x - x_0)/X_0) + B] \exp(j\omega x) \delta(y) \quad (5)$$

where Δy is the spacing of the azimuth samples in the processor or, equivalently, the pixel size in the azimuth dimension at the output detector. The spatial carrier frequency, ω , is equal to ω , divided by the acoustic velocity of the AOD. The parameters k_1 and X_0 are determined by the chirp rate and the length of the transmitted LFM radar pulse respectively.

The second AOD contains the azimuth filter signal and a sinusoidal reference which is chosen to be at the same frequency as the reference sinusoid in the first AOD. The azimuth reference filter is a LFM signal with chirp rate b_1 . The azimuth and sinusoidal reference signals have equal amplitudes to achieve maximum fringe depth in the interferometrically generated azimuth filter. As mentioned earlier, the azimuth correlation is computed by shifting either the integrating detector array or the filter signal in the AOD. Mathematically these operations are essentially equivalent. If we assume the grating shifts one pixel with respect to the detector array for each radar return, then the 1-D diffraction grating in the second AOD is expressed as:

$$P_2(x, y, n) = [\exp(j(b_1(y - n\Delta y) + 1) \exp(j\omega x) \delta(x) \quad (6)$$

The output CCD array detects and integrates the intensity of the doubly Bragg diffracted light, as shown in Fig. 8. The intensity at the detector is given by:

$$I(x, y, n) = |E_1(x, y, n)|^2 |E_2(x, y, n)|^2 \quad (7)$$

where E_1 and E_2 are the amplitude modulation envelopes of the diffracted light at the detector plane, for the first and second AODs respectively. The range focusing operation of the optics on the radar return signal in the x dimension can be shown to result in a light amplitude distribution that is a sinc function whose characteristics are determined by the space-bandwidth product (SBWP) of the range chirp. The light diffracted by the sinusoidal reference signal is collimated by the optics in the x dimension, as shown in Fig. 8. Both beams diffracted by the first AOD are smeared uniformly in the y dimension at the output plane. The light that is diffracted by the second AOD is essentially imaged

onto the output plane in the y dimension and smeared in the x dimension by the optics. The tilt of the cylindrical lens in Fig. 8 results in a scale change on the imaged azimuth reference chirp as a function of x. As can be seen in Fig. 6, this scale change is approximately linear in x when considering suitable range swaths for processing. The resulting modulation envelopes are given by:

$$E_1(x,y,n) = [AX_0 \text{sinc}(k_1 X_0 (x-x_0)) \exp(j\bar{k}_s (n-y_0/\Delta y)^2 (1+ax_0)/R_0) + B] \quad (8)$$

$$E_2(x,y,n) = [\exp(jb_s ((y-n\Delta y)(1+\beta x))^2) + 1] \quad (9)$$

From Eq. (7) the resulting intensity at the detector plane is therefore given by:

$$I(x,y,n) = [B^2 + (AX_0)^2 \text{sinc}^2(k_1 X_0 (x-x_0)) + 2ABX_0 \text{sinc}(k_1 X_0 (x-x_0)) \cos(\bar{k}_s (n-y_0/\Delta y)^2 (1+ax_0)/R_0)] \\ \times [1 + \cos(b_s ((y-n\Delta y)(1+\beta x))^2)] \quad (10)$$

The charge accumulated in each pixel of the CCD is proportional to the intensity incident on it. The total charge accumulated on the CCD after the N radar returns have been received from the point scatterer is:

$$Q_T(x,y) = \sum_{n=1}^N I(x,y,n) \quad (11)$$

Using Eq. (10) and assuming that $\beta x \ll 1$:

$$Q_T(x,y) = 2ABNX_0 \text{sinc}(k_1 X_0 (x-x_0)) \\ \times \sum_{n=1}^N \cos(\bar{k}_s (1+ax_0)(n-y_0/\Delta y)^2/R_0) \cos((b_s ((1+2\beta x)(y-n\Delta y))^2) \\ + \text{bias terms} \quad (12)$$

If b_s and β are selected (programmed) to be equal to $\bar{k}_s/R_0 \Delta y^2$ and $a/2$, respectively, then a discrete autocorrelation of the LPM signals in the azimuth direction results. Keeping only the significant terms, the final result is:

$$Q_T(x,y) \sim 2NABX_0 \text{sinc}(k_1 X_0 (x-x_0)) \text{sinc}((\bar{k}_s N/R_0 \Delta y)(1+ax_0)(y-y_0)) \\ + NB^2 + N(AX_0)^2 \text{sinc}^2(k_1 X_0 (x-x_0)) \quad (13)$$

Fig. 9 is a sketch of the charge profile which is stored on the CCD as given by Eq. (13). The first term in Eq. (13) is a 2-D sinc function whose height is proportional to the reflective strength of the point scatterer and whose center is located at coordinates (x_0, y_0) corresponding to the actual location of the scatterer on the ground. The width of the sinc function in each dimension is inversely proportional to the SBWP of the unfocused image in the corresponding dimension. Thus this term represents the high resolution focused image of the point scatterer.

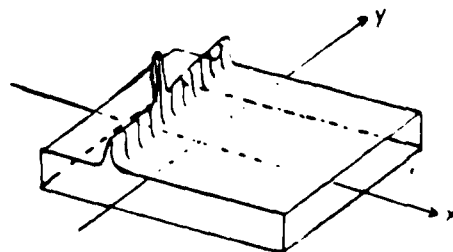


Figure 9. Focused image and bias components

The other two terms of Eq. (13) are interferometric bias components. The first is constant throughout the plane and its height is determined by the strength of the reference signal, B. The second bias term is dependent on the radar signal and results in the

formation of a ridge in the range location of the point target as shown in Fig. 9. Since each point scatterer in a target scene will have one of these bias ridges associated with it, they must be removed before a quality image can be displayed. The constant bias term should also be removed to improve the contrast of the displayed image.

Several techniques are available for removing the bias components. One technique employs a second CCD detector array on which only the signal dependent bias term is computed. The video signal output from this CCD is electronically subtracted from the video output of the original CCD to produce the bias free image. This technique is implemented using a beam splitter positioned before the second AOD in Fig. 8. The light that is split away from the main beam is range focused onto the second CCD, without azimuth processing, resulting in a charge distribution that is proportional to the signal dependent bias component of Eq. (13).

Another approach to bias removal is to employ a specially designed CCD detector array which electronically subtracts the bias on the CCD chip itself. In this technique, circuitry which is fabricated directly on the CCD substrate, is used to incrementally bleed off the bias charge as it builds up in the pixels. This approach maximizes the usable dynamic range of the CCD.

In a third bias removal technique the focused SAR image is generated on an optical spatial carrier frequency. The bias components are not modulated by the spatial carrier and are removed by electronic post-processing of the CCD video output. The carrier is generated either in the range or azimuth dimension by introducing the appropriate frequency shifts in the electronic signals. Demodulation consists of electronic filtering to detect only those components which are on the carrier, followed by envelope detection to remove the carrier frequency from the image.

Results

The real-time imaging ability of the AO/CCD system is evaluated by measuring the impulse response of the processor. Electronics are used to simulate the backscattered radar signals that would be received from a single point scatterer in the target field. The simulated radar returns are then summed to the reference signal and applied to the AOD of the processor. An isometric display is used to collect and store a frame of the CCD output and displays the focused image of the point scatterer in either a 2-D or 3-D perspective format for analysis.

An example of the undemodulated output of the processor, when displayed in 3-D perspective, is shown in Fig. 10. These data are from tests of the original processor which used a mask for the azimuth compression. The focused image of a simulated point scatterer is visible as a narrow spike which sits upon the unwanted bias ridge located in the target's range bin. The usual side-lobe structure of a focused SAR scatterer, in the azimuth dimension, is evident along the bias ridge. The CCD's fixed pattern noise as well as the nonuniformity in the illumination of the reference beam are also visible on the floor beneath the ridge. To test the resolution limitations of the CCD in the Processor, the simulated target for this data had a SBWP in excess of 100 in each dimension. Consequently, the focused image is just several pixels across and the individual pixel structure is quite evident.

Fig. 11 is an isometrically displayed image of a point scatterer when demodulation is employed. The unfocused simulated point scatterer in this example has SBWPs of approximately 40 and 100 in the azimuth and range dimensions respectively. For this data the focused image is modulated by an optical spatial carrier frequency in the azimuth dimension. An electronic filter is used to remove the bias components. This demodulation scheme is also very effective in removing the fixed pattern noises associated with the CCD and laser illumination, as is evident in Fig. 11. The dynamic range for a single simulated point scatterer was measured to be in excess of 100:1.

A maskless architecture which can electronically change its azimuth filter, like that shown in Fig. 8, has been built and its impulse response characterized. Both framing and scrolling modes of operation were demonstrated. In the framing mode a staring CCD was used, and the azimuth filter signal was electronically shifted in the AOD during the integration period to perform the azimuth compression operation. Fig. 12 is an electronically shifted version of the interferometrically generated azimuth filter shown in Fig. 6 as it would appear at one increment of the azimuth correlation. In this mode the azimuth pulse compression achieved is determined by the window size of the shifting azimuth signal and not by the CCD aperture. A simulated point scatterer with a SBWP of 40 in the azimuth dimension was successfully focused.

Fig. 13 is an image of a simulated point scatterer focused with the scrolling mode of

operation. In this implementation the azimuth compression is achieved by operating the CCD detector array in the shift-and-add mode. The azimuth reference signal that is loaded into the second AOD is the same (and not shifted in time) for each radar pulse. For this experiment the simulated scatterer had an azimuth phase history which matched that of a range bin near the center of the filter shown in Fig. 6. The focused image of Fig. 13 was generated using the azimuth carrier demodulation scheme that has been described previously. The pulse compression achieved is approximately 35 and 75 for the azimuth and range dimensions respectively, which accounts for the oblong shape of the focused spot.

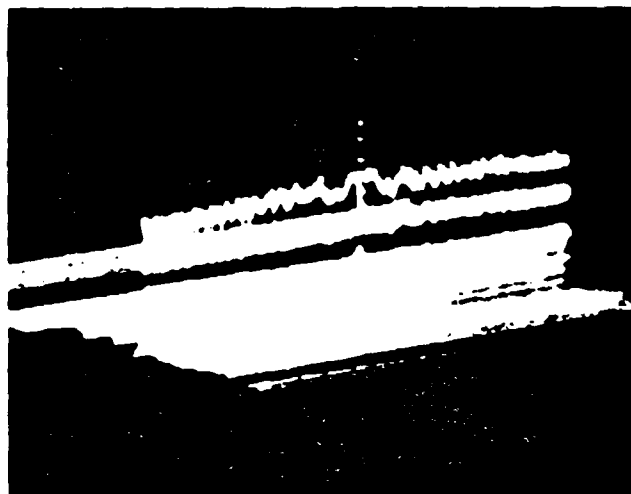


Figure 10. Example of undemodulated output.

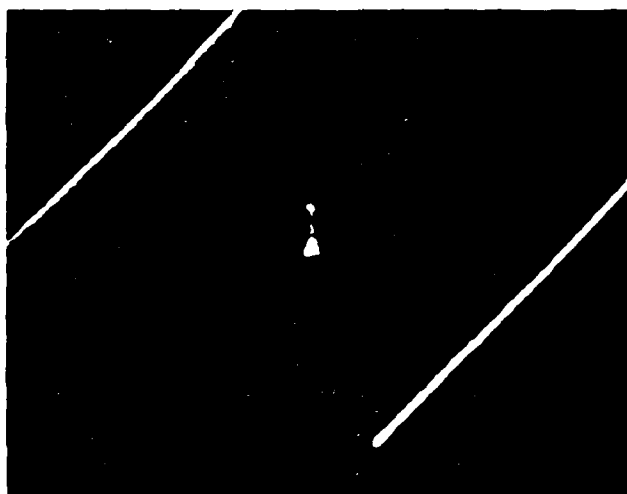


Figure 11. Example of demodulated output.



Figure 12. Electronically shifted azimuth reference function.

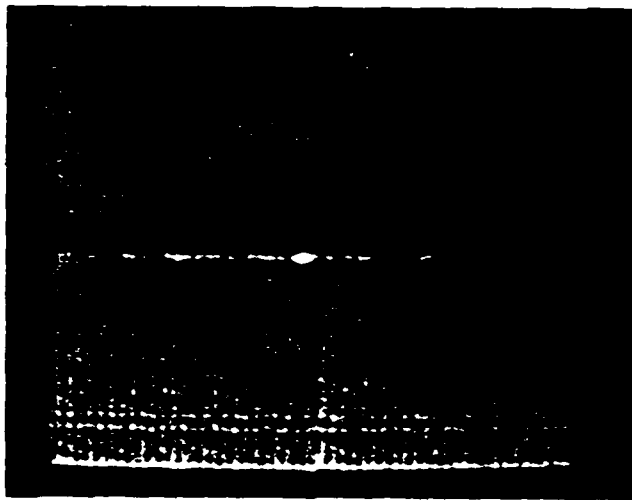


Figure 13. Simulated point scatterer focused in real time with the programmable (mask-less) architecture.

Acknowledgements

The work reported in this paper is supported by the Air Force Office of Scientific Research, General Dynamics, and NASA. The authors wish to thank K. Wagner and J. Yu for their assistance.

References

1. H. L. McCord, "The Equivalence Among Three Approaches to Deriving Synthetic Array Patterns and Analyzing Processing Techniques," IRE Transactions on Military Electronics, Vol. MIL-6, No. 1, pp. 116-119, January 1962.
2. L. J. Cutrona, et al, "On The Application of Coherent Optical Processing Techniques to Synthetic Aperture Radar," Proc. IEEE, 54, 1826 (1966).
3. D. Psaltis and K. Wagner, "Real-time Optical SAR Processor," Opt. Eng., Vol. 21, p. 822, 1982.
4. D. Psaltis, M. Baney, and K. Wagner, "Real-Time Synthetic Aperture Radar Processing," Proc. NASA Conf. on Optical Information Processing, (Hampton, Va., 1983).
5. M. Baney, K. Wagner, and D. Psaltis, "Programmable Real-Time Acousto-optic/CCD SAR Processor," SPIE Vol. 495, August, 1984.
6. K. Wagner and D. Psaltis, private correspondence

TIME AND SPACE INTEGRATING ACOUSTO-OPTIC FOLDED SPECTRUM PROCESSING FOR SETI

Kelvin Wagner and Demetri Psaltis

Department of Electrical Engineering
California Institute of Technology
Pasadena, Ca. 91125

Abstract

Time and space integrating folded spectrum techniques utilizing acousto-optic devices (AOD) as 1-D input transducers are investigated for a potential application as wideband, high resolution, large processing gain spectrum analyzers in the search for extra-terrestrial intelligence (SETI) program. The space integrating Fourier transform performed by a lens channelizes the coarse spectral components diffracted from an AOD onto an array of time integrating narrowband fine resolution spectrum analyzers. The pulsing action of a laser diode samples the interferometrically detected output, aliasing the fine resolution components to baseband, as required for the subsequent CCD processing. The raster scan mechanism incorporated into the readout of the CCD detector array is used to unfold the 2D transform, reproducing the desired high resolution Fourier transform of the input signal.

Introduction

The Search for Extra-Terrestrial Intelligence (SETI) is an ambitious program to examine 773 of the nearest solar type stars as well as the entire celestial sphere for any evidence of intelligent microwave activity, either beamed at us or leaked from a civilization's internal communications¹. The interstellar distances are so large and the resulting $1/R^2$ attenuation is so great that the receiver sensitivity to a signal buried in cosmic background noise must be as large as possible. Without a priori knowledge as to the transmitted signal structure ideal matched filters cannot be constructed, however a good assumption is that a transmitted beacon would have a narrow band highly coherent structure in a sea of broadband white noise. Thus a narrow bandpass filter, with resolution as small as 1-100 Hz, could achieve enough processing gain to detect the presence of an intelligent CW signal. Since the preferred frequency for interstellar microwave communication could conceivably lie anywhere from .1-10-100 GHz, then from 10^8 to 10^{11} frequency measurements must be made for each resolvable spot on the celestial sphere observed by the radio telescope. If these measurements were performed sequentially with a swept local oscillator tuned bandpass filter then as much as a million years of observation time could be required. A much more practical approach would be to build a wideband, high resolution spectrum analyzer with a million or more parallel channels of band pass filters thereby cutting down observation time to on the order of a year. Currently two approaches are being pursued, a digital multi-channel spectrum analyzer (MCSA)² with a tree configuration of DFT and FFT subunits, and an analog acousto-optic folded spectrum processor as reported in this paper.

The advanced state of the art of wideband acousto-optic (AO) Bragg cells make them ideal candidates for entering 1-D data into an optical signal processing system, and the Fourier transforming property of a lens makes the implementation of real-time spectrum analyzers with 1000 resolvable frequencies routine. In this paper we explore some possible approaches to performing narrow band time integrating (TI) spectrum analysis on each of the 10^3 channels of a space integrating (SI) AO spectrum analyzer, thereby obtaining a high resolution folded spectrum through time and space integration (TSI).

Folded Spectrum Processing

The concept of the Folded Spectrum was introduced by Thomas², who showed that the 2-D Fourier transform of a raster scanned signal was an orthogonally raster scanned version of the long 1-D spectrum of the original signal. This is schematically illustrated in Fig. 1. The folded spectrum is ideal for an optical signal processing system since it utilizes the 2 available spatial coordinates to perform a high resolution spectrum analysis that would be impossible in a 1-D system due to the limitations in space bandwidth product of 1-D devices. This processing operation can be performed in real time by raster scanning a very long input signal $s(t)$ onto a 2-D spatial light modulator (SLM), and using the 2-D space integrating Fourier transform operation performed by a spherical lens. After the complete 2-D raster is formed, coherent light is passed through the SLM and the resulting amplitude distribution $f(x,n)$ is Fourier transformed by the lens. The power spectrum $|F(u_c, v_f)|^2$ is detected by a 2-D charge coupled device (CCD) detector array at the back focal plane. The

orthogonally oriented raster readout of the CCD turns the 2-D power spectrum back into a high resolution 1-D power spectrum $|S(f)|^2$, at a rate of up to 60 such transformations each second. However, the availability of highly mature CCD devices with up to 2000 x 2000 pixels can not be fully utilized because of the lower resolution and dynamic range of currently available 2-D SLMs. Thus an architecture for performing the high resolution folded spectrum operation without the need for 2-D SLMs is highly desirable.

Turpin³ and Kellman⁴ have suggested such an approach utilizing acousto-optic devices (AODs) in an interferometric time integrating configuration. In this approach a 2-D array of local oscillators with coarse and fine frequencies in x and y is produced by orthogonal AODs to which fast and slow chirps are applied respectively. When a frequency component of the input signal matches that of a particular local oscillator, a baseband heterodyne component is generated, and a time integrated peak is created at that location, proportional in amplitude to the amplitude of the corresponding Fourier component. Since this approach involves interferometric detection, a bias term will be generated due to the local oscillator reference and a signal dependent bias will occur uniformly throughout the detector array. These components can be separated by placing the desired heterodyne term on a spatial carrier. However in a white noise environment as envisioned for the SETI search with a weak narrow band CW peak, the signal dependent bias term will produce a large bias because all its frequency components are incoherently detected at each CCD pixel, thereby using up the available CCD dynamic range. Thus the 2-D time integrating technique for folded spectrum processing suffers in a multisignal environment from a large signal dependent bias term, and places too large a requirement on CCD dynamic range.

A hybrid approach that utilizes the 1-D space integrating Fourier transformation property of a lens in one dimension followed by an interferometric time integrating fine resolution spectral analysis can combine the best features of the 2-D SI and 2-D TI approaches to folded spectrum processing. This approach utilizes a standard space integrating acousto-optic spectrum analyzer (Fig. 3) to perform the coarse spectral analysis without introducing a signal dependent bias term. The number of resolvable coarse frequency bins is limited by the time bandwidth product of the AOD and the input apodization to approximately $T_B = 1000$. The doppler frequency shift associated with acousto-optic Bragg diffraction produces a linear spatial dependence of the temporal frequency at the output spectrum that must be converted to baseband at all positions for further time integrating processing on a CCD. Bader^{5,6} proposed a way to do this with a spatially distributed local oscillator (DLO) reference wave which is produced with an auxiliary AOD that has a repetitive wide band signal applied to its transducer. Interferometric detection of the signal and reference Fourier spectras produces narrowband baseband signal components at each output location. These narrowband signals can be used to produce a time integrating fine frequency analysis in the orthogonal direction on a 2-D CCD through the use of the chirp transform algorithm by premultiplying each narrowband signal with a slow chirp and then convolving with a matched chirp using another AOD. This type of system has a less severe signal dependent bias build up problem because only signals within a single coarse frequency bin contribute any bias to that bin. In this paper we will present an alternative approach to time and space integrating folded spectrum processing.

This method, which utilizes pulsed laser diodes, was originally proposed by Psaltis^{7,8}, and is illustrated schematically in Fig. 2. A very long wideband signal is to be optically processed utilizing an acousto-optic device as the input transducer. However, since the limited time-bandwidth product of the AOD will not allow the entire signal to be entered into the AOD simultaneously, we will successively slide overlapping portions of the long signal into the AOD and freeze the acoustic motion by pulsing the illuminating laser diode with a very narrow pulse. For each laser diode pulse we have a 1-D spatial representation of part of the input signal as an optically coherent wavefront, and on successive pulses we will transduce successive portions of the input signal into the optical system. Thus we have transformed the long signal into a 2-D raster of space (x), and time or pulse number (n). The transformation performed by the AOD illuminated with a repetitively pulsed laser diode is given by

$$f(x,n) = \sum_{n=0}^{N-1} s(t-x/v) \text{rect}(x/A) \delta(t-nT). \quad (1)$$

In this expression v is the acoustic velocity of the travelling wave AOD, A is the AOD width, and T is the laser diode pulse repetition interval. If $A > vT$ the raster is oversampled, which means that successive raster lines are overlapping. We perform the desired long 1-D Fourier transform on the input signal $s(t)$, with a 2-D transform on its oversampled 2-D raster representation $f(x,n)$. Because of the separability of the multidimensional Fourier kernel, a 2-D transform can be performed as N 1-D transforms along one axis, followed by N 1-D transforms along the orthogonal axis. This cascaded

2-D Space Integrating Folded Spectrum

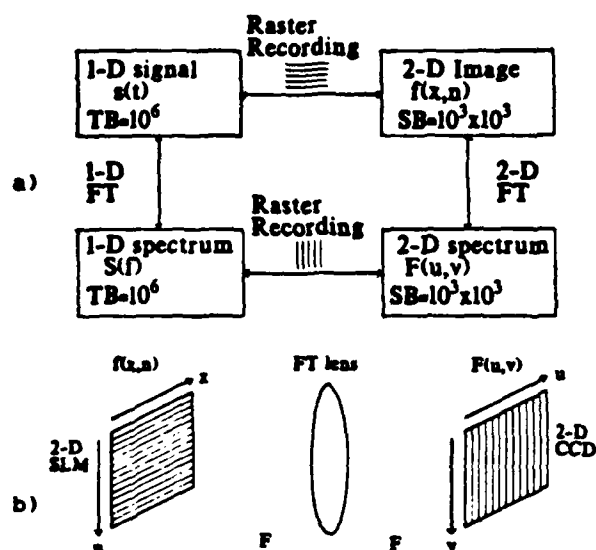


Fig. 1. a) Relationship between a long 1-D signal, its high resolution spectrum, the folded spectrum, and the signal's raster representation. b) A space integrating folded spectrum processor.

TSI Processing of Long 1-D Signals as 2-D Rasters of Space and Time

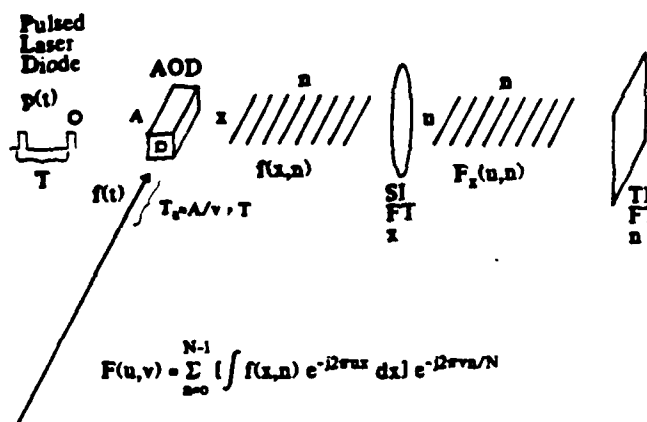


Fig. 2. Time and Space integrating folded spectrum processor showing the conversion to a space-time raster, spatial Fourier transformation followed by array of temporal Fourier transforms.

system can be implemented by performing a spatial Fourier transform with a lens for each of the N laser diode pulses, producing an intermediate data set $F_x(u,n)$, which is also a space-time raster. This is followed by an array of time integrating Fourier transform modules that utilize the second spatial dimension (y) to perform the desired accumulation of partial products. After a succession of N laser diode pulses the full folded spectrum is produced in a 2-D format on a CCD detector array. The 2-D spectrum can be very simply unfolded back into the desired long 1-D spectrum by utilizing the raster readout mechanism incorporated in the CCD itself. The resulting system utilizes the mature component technologies of acousto-optics, charge coupled devices and laser diodes to produce a valuable real-time signal processing module for high resolution spectral analysis.

Space Integrating Spectrum Analyzer with a Pulsed Laser Diode

The operation of a space integrating (SI) acousto-optic spectrum analyzer is explained with reference to Fig. 3. A collimated coherent optical plane wave of wavelength λ is incident upon the AOD at the Bragg angle, $\theta_B = -\sin^{-1} \lambda/2 \cdot \omega/v$, in order to maximize the optical power coupled into the first order diffracted beam. In this expression f_c is the transducer center frequency, $\omega = 2\pi f_c$ is the midband acoustic wavelength, and v is the acoustic velocity. A bandlimited signal is mixed to the AOD center frequency f_c , producing an octave bandwidth signal $s(t)$, that is amplified and applied to the piezoelectric transducer, thereby launching an acoustic wave replica of $s(t)$ into the photoelastic medium. The acoustic wave creates a travelling wave volume index perturbation through the photoelastic effect, which Bragg diffracts some of the incident light into the first order diffracted optical replica of a single sideband version of the portion of the signal $s(t)$ that is currently within the AOD aperture A . In the frequency domain this can be viewed as each Fourier component $S(f)$ of the signal $s(t)$ within the AOD bandwidth $B=2f_c/\lambda$, producing a diffracted plane wave of aperture A propagating at an angle $\theta_D = \lambda f/2v$ with respect to the optic axis, doppler shifted by the frequency f due to the acoustic wave motion. This diffracted field propagates through a distance F and is incident on a Fourier Transforming lens of focal length F , so that at a distance F behind the lens the spatial Fourier transform of the signal within the AOD is formed. At the back focal plane the field observed is

$$\begin{aligned}
 P(x', t) &= \int_A s(t - x/v) e^{-j2\pi x x' / \lambda F} dx e^{j\omega t} \\
 &= \int_A \left[\int_B S(f) e^{j2\pi f(t - x/v)} df \right] e^{-j2\pi x x' / \lambda F} dx e^{j\omega t} \\
 &= \int S(f) e^{j2\pi f t} A \operatorname{sinc}[A(x' / \lambda F - f/v)] df e^{j\omega t} \\
 &= [v S(x'v / \lambda F) e^{j2\pi v x' t / \lambda F} e^{j\omega t}] * A \operatorname{sinc}(Ax' / \lambda F). \quad (2)
 \end{aligned}$$

In this expression x is the spatial coordinate along the axis of acoustic propagation, x' is the spatial coordinate at the Fourier plane, and $\omega = 2\pi c/\lambda$ is the optical radian frequency. The symbol $*$ represents the convolution operation. We have neglected effects of input apodization, acoustic attenuation, anisotropic diffraction and band limitation, as well as other higher order effects. The exact doppler temporal phase history of each Fourier component is reproduced at the back focal plane, but the spatial representation of the spectrum is blurred by the Fourier transform of the finite aperture function of width A thereby limiting the effective frequency resolution to $\delta f = 1/T_a = v/A$, where T_a is the acoustic transit time across the aperture. This shows that the number of independent analog samples stored within the AOD at any particular instant, which is given by the time bandwidth product $T_a B$, also gives the effective number of independent frequencies that can be determined in the spatial Fourier transform plane. The actual frequency resolution is limited by the AOD finite aperture sinc convolved with the transforms of the optical and acoustic apodization functions, in order to produce an overall blur spot $Q(x'/\lambda F)$ which yields somewhat worse frequency resolution than $1/T_a$. Normal operation of an acousto-optic spectrum analyzer involves the square law detection process on a 1-D array of integrating photosensors, which incoherently accumulate charge for a time T' in order to improve the signal to noise ratio.

$$P(x') = \int_0^{T'} |P(x', t)|^2 dt = \int_0^{T'} |S(x'v/\lambda F) * Q(x'/\lambda F)|^2 dt \quad (3)$$

The output of this type of spectrum analyzer is a power spectrum of the signal $s(t)$ with a resolution limited by the blur spot $Q(u)$, with a normalized spatial frequency variable $u = x'/\lambda F$. For a single sinusoidal input of frequency f' input to the spectrum analyzer, $s(t) = 2a \cos(2\pi f' t)$, with a single sided temporal Fourier spectrum given by $S(f) = a \delta(f - f')$, the time integrated output of the spectrum analyzer is given by

$$P(u) = T' |a|^2 Q^2(u - f'/v). \quad (4)$$

A heterodyne detection technique can be employed by the addition of a simple plane wave reference beam. The instantaneous output of such an interferometric spectrum analyzer has a linear spatial dependence of the doppler induced temporal frequency that is given by

$$\begin{aligned}
 P(u, t) &= |[S(uv) e^{j2\pi v u t} e^{j\omega t}] * Q(u) + r e^{j\omega t}|^2 \\
 &= |S(uv) * Q(u)|^2 + |r|^2 + 2r [S(uv)] \cos(2\pi v u t + \phi(uv)) * Q(u). \quad (5)
 \end{aligned}$$

Where the amplitude and phase of the complex spectrum $S(f) = |S(f)| e^{j\phi(f)}$ are reproduced as temporal modulations of each resolved frequency bin which is spatially blurred due to the apodization. For the case of the single sinusoidal input tone at a frequency f' , with amplitude $|a|$ and phase ϕ , the heterodyne detected instantaneous output is

$$\begin{aligned}
 P(u, t) &= |a Q(u - f'/v) e^{j2\pi f' t} e^{j\omega t} + r e^{j\omega t}|^2 \\
 &= |a|^2 Q^2(u - f'/v) + |r|^2 + 2|a|r Q(u - f'/v) \cos(2\pi f' t + \phi). \quad (6)
 \end{aligned}$$

The temporal modulation on the last term thus reproduces the input sinusoid in both amplitude and phase, and could be used as the input to a fine resolution time integrating processor if we first heterodyne it to baseband. A repetitively pulsed light source is a simple and effective means of accomplishing this as long as the pulse width τ is less than the inverse of the bandwidth of $s(t)$, and much greater than the optical period of oscillation. When these conditions are satisfied we can consider the laser diode pulsed modulation to be approximated by a series of delta functions separated by a time T .

$$p(t) = \sum_n s(t - nT). \quad (7)$$

This series of delta functions will act as sampling impulses on the interferometrically detected temporal modulation present at the output plane. It is convenient to consider the input temporal frequency to be broken up into a coarse and fine part by choosing the closest harmonic of the laser diode pulse repetition frequency as the coarse term and the offset from this frequency as the bandlimited fine resolution component. Thus we substitute $f' = k/T + \delta$ into equation 6, where $k = \text{mod}_1/T$ is an integer and $\delta = \text{res}_1/T \ll 1/T$. For the case of repetitively pulsed laser illumination the heterodyne output of the AO spectrum analyzer for a single applied sinusoidal tone, is given by

$$P(u, t) = \{ |a|^2 Q^2(u - (k/T + \delta)/v) + |r|^2 + 2|a|r Q(u - (k/T + \delta)/v) \cos(2\pi(k/T + \delta)t + \theta) \} p(t). \quad (8)$$

The fine frequency component δ will not move the blur spot $Q(u)$ by a significant amount and can be dropped from within $Q(u)$. The coarse frequency component k/T will be sampled by the laser diode pulses at times $t = nT$ resulting in an argument within the cosine function of $nk2\pi$, with n and k both integers, so that this term can be dropped leaving only the fine frequency modulation. Thus for the n th laser diode pulse at time $t = nT$ the sampled output can be approximated as

$$P(u, nT) = |a|^2 Q^2(u - k/Tv) + |r|^2 + 2|a|r Q(u - k/Tv) \cos(2\pi\delta nT + \theta). \quad (9)$$

Thus for a single sinusoidal input the interferometric output of the space integrating spectrum analyzer illuminated with a series of delta functions can be seen to consist of a signal dependent bias term located at the position corresponding to the coarse frequency component, a uniform reference bias term and a sampled interferometrically detected spot located at the coarse frequency locus and oscillating in time at the fine frequency δ , with phase θ . This effect of the sampling pulse train is usually referred to as aliasing and is considered undesirable. It indicates that the sampling rate is too low for the given signal bandwidth, and it results in different parts of the signal spectrum folding over on top of each other thereby producing invalid signal components. However in this case, the coarse frequency channelization of the SI spectrum analyzer effectively separates in space the various frequency components that could alias with each other. It is important to make sure that the sampling rate is fast enough to avoid aliasing of the bandlimited signals within each coarse frequency blur spot $Q(u)$. Since the bandwidth of the signal information within each blur spot is increased over $1/T_a$ due to the apodization effects, we will assume that an appropriate bandwidth would be $2/T_a$. To adequately sample these bandlimited signal components the Nyquist criterion indicates that we should sample at least twice as fast as the highest frequency signal component present, thus the minimum sampling rate should be $1/T > 4/T_a$. This indicates that the laser diode should be pulsed at least 4 times as an acoustic signal transits across the laser beam aperture illuminating the AOD in order to produce the appropriately oversampled space-time raster.

Acousto-Optic Spectrum Analyzer

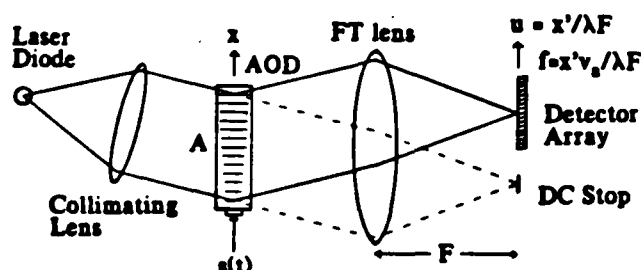


Fig. 3. The space integrating acousto-optic spectrum analyzer.

Time Integrating DFT Spectrum Analyzer

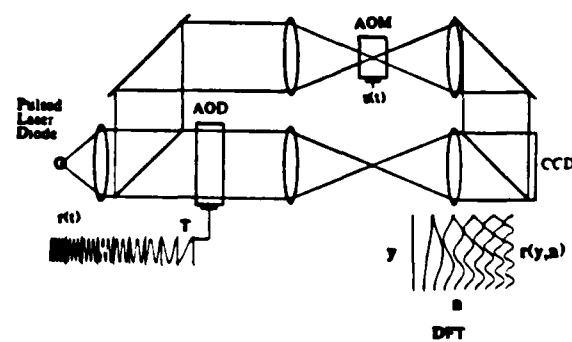


Fig. 4. Interferometric time integrating DFT narrowband spectrum analyzer with phase locked stepped frequency reference.

Time Integrating DFT Based Fine Resolution Processor

A narrowband fine resolution spectrum analysis operation requires the coherent accumulation of data samples for a period at least as long as the inverse spectral

resolution. For this reason a SI AO spectrum analyzer is limited in resolution to the inverse aperture time of the AOD. In order to obtain a higher resolution spectrum analysis with AODs a time integrating (TI) approach must be utilized. Coherent accumulations can be carried out on an integrating charge coupled device (CCD) detector array as long as the thermally generated detector dark current and optical bias contributions do not saturate the desired signal components. With typical CCD device defect densities, thermal generation rates and signal well capacities it should be possible to integrate for a full second, obtaining 1 Hz resolution capabilities, and with cooled detectors it would be possible to integrate even longer. For the SETI program these resolution capabilities should be more than sufficient since spectral coherence of better than 1 Hz would be very difficult to obtain after propagating through the interstellar medium and being doppler shifted by both transmitter and receiver motions.

The required operation that needs to be performed on the output of each coarse frequency bin of the interferometric SI spectrum analyzer with a pulsed LD reported in the previous section is a narrowband discrete Fourier transform (DFT). The DFT is the discrete version of the Fourier transform as is appropriate for sampled data streams such as that produced by the interferometric spectrum analyzer illuminated by a pulsed laser diode. If we examine the third term in Eq. 9 we can see that for each input sinusoid we have a sampled fine frequency temporal oscillation $s(n)$ that we want to perform a DFT on. The DFT is defined for such a sampled input signal $s(n)$ as

$$S(m) = \sum_n^N s(n) e^{-j2\pi nm/N}. \quad (10)$$

This operation is a vector matrix multiplication, where the sampled input signal $s(n)$ is the input vector, the DFT kernel is the matrix, and $S(m)$ is the output vector. A possible time integrating interferometric optical implementation using a pulsed laser diode is shown schematically in Fig. 4. The light from a repetitively pulsed laser diode is split into two by a beam splitter. One arm illuminates an acousto-optic modulator that has a narrowband signal $s(t)$ applied to its transducer sampled by the narrow pulse of the laser diode producing a temporal modulation $s(nT)$. In the other arm, a reference AOD with aperture T , equal to the LD pulse repetition interval, is schlieren imaged onto the output CCD array where it is interferometrically combined with the spatially uniform signal beam producing a sinusoidal spatial fringe pattern. With each laser diode pulse a new column of the DFT matrix is applied to the reference AOD transducer, and propagates across the aperture to align with the CCD at which time its motion is frozen by the LD pulse thereby imaging that column onto the CCD. Thus the reference signal is analogous to a raster scanned version of the DFT matrix, except that along the columns a continuous representation is used which is later sampled by the spatially discrete CCD pixel structure. Since the successive columns of the DFT matrix represent sinusoids of successively higher frequencies the reference waveform is a stepped frequency chirp, with the constraint that each new frequency phase resets to zero phase each T . The reference signal starts at frequency f and frequency steps by Δf each time T which is analytically represented as

$$r(t) = \sum_{n=0}^{N-1} \text{rect}\left[\frac{t-nT}{T}\right] \cos[2\pi(f+n\Delta f)(t-nT)]. \quad (11)$$

This signal propagates in the y direction at a velocity v , and aligns with the AOD and CCD apertures at the times of the laser diode pulses as given in Eq. 7. The spatial modulation diffracted by the AOD on the n th pulse is given by

$$\begin{aligned} r(y, n) &= r(t-y/v) \text{rect}[y/vT] \delta(t-nT) \\ &= \text{rect}[y/vT] e^{j2\pi(f+n\Delta f)y/v}. \end{aligned} \quad (12)$$

Which is seen to be a finite aperture plane wave with an angular spatial frequency that is stepped linearly with the pulse number n . The interferometrically detected intensity pattern on the n th LD pulse is found by combining weighted signal and reference beams. The signal beam propagates at an angle θ with respect to the starting frequency of the reference, resulting in a spatial carrier of spatial frequency $a = \sin\theta/\lambda$.

$$\begin{aligned} I(y, n) &= |G s(nT) e^{j2\pi(a+f/v)y} + b r(y, n)|^2 \\ &= |G|^2 + |b|^2 + 2 G b s(nT) \cos[2\pi(n\Delta f y/v + ay)] \end{aligned} \quad (13)$$

By time integrating over N successive LD pulses we obtain a discrete summation of the interferometrically detected charge distributions due to each pulse.

$$I(y) = N (|G|^2 + |b|^2) + 2 G b \sum_{n=0}^{N-1} s(nT) \cos[2\pi(n\Delta f y/v + \phi y)] \quad (14)$$

This can be recognized as the discrete cosine transform of the sequence $s(nT)$, with an analysis bandwidth of Δf , on a spatial carrier of frequency $\gamma = \Delta f + (N-1)\Delta f/2v$. The carrier allows a bandpass filter operation to remove the bias terms, and the full complex DFT can be reconstructed by measuring the phase angle of the carrier for each resolvable frequency component. The CCD is composed of an array of pixels that must have sufficient resolution to adequately sample the spatial fringe structure under the DFT transform. For a pixel size Δy , the number of pixels per cycle of the fringe structure is $1/\gamma\Delta y$, and should be adjusted to 4 to 8 pixels per cycle. The total number of CCD pixels required is the product of the DFT space bandwidth product with the number of pixels per frequency resolution element, which must be at least one cycle, thus $M = N\Delta f/\gamma\Delta y > 4N\Delta f$.

TSI Folded Spectrum Based on TI DFT

The characteristics of the TI DFT processor described in the previous section are complementary to those of the SI AO spectrum analyzer. The TI DFT processor is a narrow band processor with analysis bandwidth $\Delta f < 1/2T$, and high resolution of $1/NT$. The SI AO spectrum analyzer is a broadband processor with bandwidth B , and resolution $\sim 2/T$, which may also be limited by the CCD pixel size which has a frequency width of $v\Delta x/\Delta f$. Since the TI DFT processor is a 1-D system, we can spatially multiplex an array of such processors in the orthogonal dimension by using a 2-D CCD detector array. Since each of the TI DFT processors in the array needs to perform the same narrowband spectrum analysis operation, and each signal is brought down to baseband by the pulsed laser diode, then they all can use the same reference AOD. One way that we can configure a 2-D acousto-optic folded spectrum processor is as an additive Mach-Zehnder interferometer with a SI AO spectrum analyzer to perform coarse frequency analysis in the x dimension in one arm of

Additive Mach-Zehnder Interferometer
For Acousto-Optic Folded Spectrum

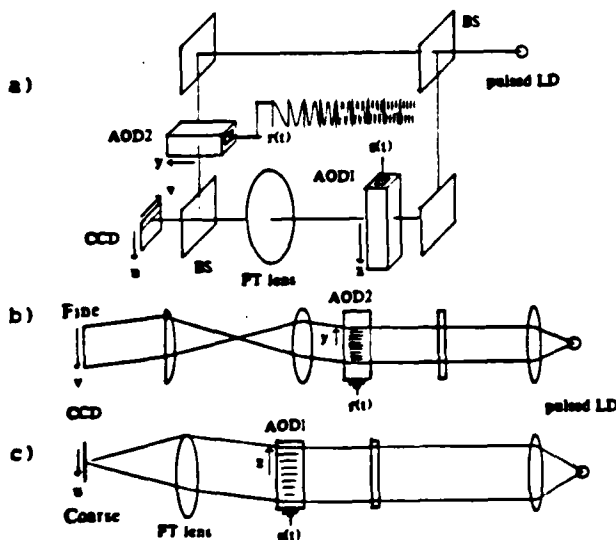


Fig. 5. a) Mach-Zehnder interferometer for performing TSI folded spectrum. b) Cross-section through TI fine resolution spectrum analyzer arm. c) Cross-section through SI coarse resolution spectrum analyzer arm.

Single Tone Output

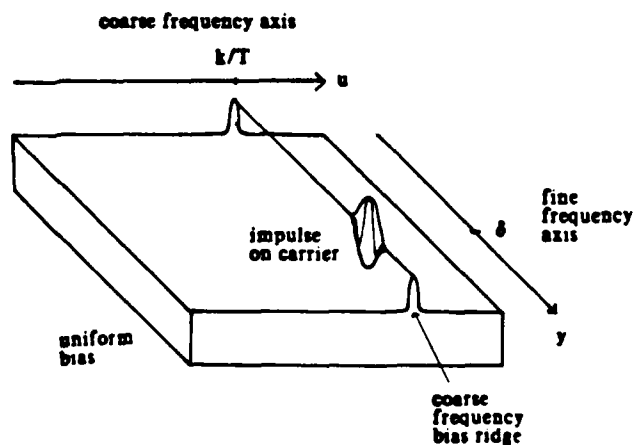


Fig. 6. Impulse response for TSI folded spectrum processor for a single tone input at frequency $f' = k/T + \delta$, showing uniform bias, signal dependent bias, and impulse riding on carrier.

the interferometer, and a spatially multiplexed array of TI DPT processors in the other arm in order to perform fine frequency analysis in the y dimension on each coarse frequency bin. The fine frequency analysis bandwidth Δf should be equal to the coarse frequency resolution per pixel in order to produce a full folded spectrum without any gaps.

The description and analysis of the TSI AO folded spectrum processor is made with reference to Fig. 5, which schematically illustrates the Mach-Zehnder architecture in the upper half, and shows cross-sections of the fine and coarse arms in the bottom half of the figure. The timing of the system is controlled by a 10 MHz stable crystal oscillator that is digitally divided by 100 to produce a stable 100 KHz pulse train used as the laser diode trigger. The laser diode is biased just below threshold and pulsed with a 50 nsec pulse every 10 μ sec. The anamorphic Gaussian beam profile of the laser diode is collimated by the spherical lens and oriented with the long axis along the x direction in order to optimally illuminate the entire SI AOD aperture. For the initial single tone experiments reported in this paper the beam splitter reflects only about 10% of the light to the SI arm of the interferometer, because the processing gain associated with the spatial integration performed by the Fourier lens collects enough light to make this arm brighter at the output plane, so the optical efficiency can be increased by directing more light to the TI arm. For an actual SETI signal environment of broadband noise, the beam ratio can best be balanced through the use of a 50% beam splitter. In the SI arm a cylindrical lens with power in y focuses the light to a narrow slit positioned to coincide with the center of the diffracting acoustic beam, and incident on the AOD at the Bragg angle. The AODs used in this experiment are slow shear mode TeO_2 devices with an aperture $T_A = 70 \mu\text{seconds}$ and a bandwidth $B = 40 \text{ MHz}$. The signal $s(t)$ applied to the transducer of AOD1 produces a diffracted component which is Fourier transformed in x by the following spherical lens, which simultaneously recollimates in y. Thus each frequency component of $s(t)$ produces a uniform slit of light at the output plane, and the position of this slit varies linearly with the input coarse frequency. In the TI arm of the interferometer a cylindrical lens with power in x is used to focus the light down to a narrow slit along y which is incident on AOD2 at the Bragg angle. Only a 10 μ second portion of the AOD needs to be uniformly illuminated with the narrow dimension of the LD elliptical beam profile. The reference signal is generated by incrementing a 10 bit digital counter with each LD pulse, and driving a 10 bit digital-to-analog converter in order to make a 1024 level staircase lasting 10.24 msec. This signal is then applied to a phase resettable voltage controlled oscillator whose phase is reset each LD pulse producing the waveform $r(t)$ of Eq. 11, which is amplified and applied to AOD2. The diffracted signal from AOD2 is Fourier transformed by the spherical lens which recollimates in the x dimension and allows a bandpass filter to be placed in the transform plane. The filtered light is retransformed in y by the following cylinder, producing an image of the signal in AOD2 with the appropriate scale so that one 10 μ second sinusoid fills the entire CCD aperture B in the y dimension, and is uniformly spread out in the x dimension. The final beamsplitter recombines the two beams at an adjustable angle, and the CCD time integrates the interferometric product between the two waves for a single video frame time of 16.7 msec. The CCD vertical synchronization signal is used to reset the 10 bit digital counter and begin another folded spectrum calculation by time integrating the next 1024 laser diode pulses.

The operation of the processor can be described analytically by combining the results of the previous two sections. The optical field incident on the CCD during the nth pulse is the sum of the reference wave acting as a linear spatial frequency modulation in the y dimension, with the spatially Fourier transformed signal within the aperture of AOD1. The photogenerated charge profile detected by the CCD on each LD pulse is the modulus squared of the field incident during that pulse. Thus the time integrated interferometrically generated charge distribution recorded on the CCD after a full frame of processing is given by Eq. 15.

$$\begin{aligned}
 I(u, y) &= \sum_{n=0}^{N-1} \left| b e^{j2\pi(f+n\Delta f)y/v} + G e^{j2\pi(a+f/v)y} \int_A f(x, n) e^{-j2\pi ux} dx \right|^2 \\
 &= N |b|^2 + N |G \int_A f(x, n) e^{-j2\pi ux} dx|^2 \\
 &\quad + 2 b G \sum_{n=0}^{N-1} \left[\int_A f(x, n) e^{-j2\pi ux} dx \right] \cos[2\pi(n\Delta f y/v + ay)]
 \end{aligned} \tag{15}$$

The phase reset reference wave starts at frequency f and steps through N discrete frequencies separated by Δf , and the angle of incidence in the y direction θ can be adjusted to make the resulting fringe structure at a desired spatial frequency $\gamma = a + (N-1)\Delta f/2v$. The output pattern consists of a uniform bias term due to the reference

beam, a signal dependent bias term with no variation in the y dimension, and the last term which is recognized as the 2-D Fourier transform of $f(x,n)$, which is the desired folded spectrum of the long 1-D signal $s(t)$. The folded spectrum term is riding on a spatial carrier of spatial frequency γ in the y dimension so that it can be demodulated from the bias terms by an appropriate electronic bandpass filter. For the case of a single sinusoidal signal of frequency $f' = k/T + \delta$, amplitude $|a|$, and phase α , the output becomes

$$I(u,y) = \sum_{n=0}^{N-1} |b| e^{j2\pi(f+n\Delta f)y/v} + G e^{j2\pi(\alpha+f/v)y} a Q(u-k/Tv) e^{j2\pi\delta nT} |^2$$

$$= N |b|^2 + N G^2 |a|^2 Q^2(u-k/Tv)$$

$$+ 2 G |a| |b| Q(u-k/Tv) N \text{sinc}[NT(\delta - \Delta f y/H)] \cos(2\pi\gamma y + \alpha) \quad (16)$$

This pattern is the single tone impulse response of the system and is shown schematically in Fig. 6. It consists of a uniform bias term, a signal dependent bias ridge at the coarse frequency location $u = k/Tv$, and the folded spectrum impulse of amplitude $|a|$ riding on top of the ridge at fine frequency position $y = \delta H/\Delta f$. The impulse is on a spatial carrier of frequency γ in the y dimension, with phase angle α which will vary on successive frames according to its super fine frequency $\Delta = \text{res}_y/NT\delta$, thus processing of successive frames can achieve even finer resolution. The resolution of Eq. 16 is given by the width of the sinc function and is limited by the integration time to $1/NT$, however the CCD spatial frequency response will apodize the DFT and decrease the achievable resolution. The fine frequency analysis bandwidth is seen to be Δf as y is allowed to vary over the CCD height H .

Experimental results from this processor are shown in Fig. 7 for a fine frequency analysis bandwidth $\Delta f = 4$ KHz. The processor was operated with a coarse frequency resolution per pixel of 80 KHz, but due to limitations of the electronics the fine frequency analysis

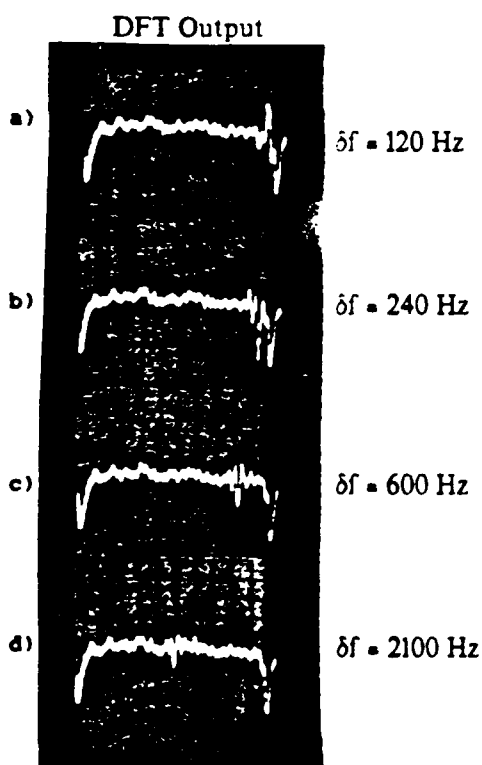


Fig. 7 Experimental results for the TSI folded spectrum processor of Fig. 5, showing crosssections through the coarse frequency ridge for fine frequencies of a) 120 Hz b) 240 Hz c) 600 Hz d) 2100 Hz.

Time Integrating Chirp Transform on TDI CCD

$$F(u,v) = \sum_{n=0}^{N-1} f(x,n) e^{-j2\pi v x} dx + r \exp[-j\pi(nT)^2 v^2] \cos(by^2) + \delta(nT - y/V)$$

SI reference chirp chirp mask TDI CCD

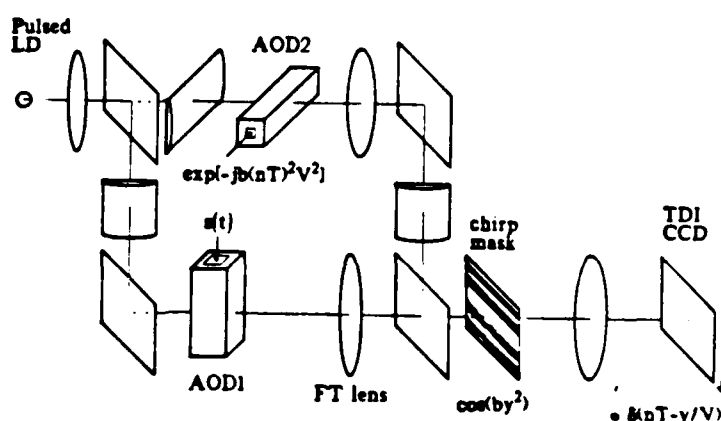


Fig. 8. Architecture for performing TSI folded spectrum using a TDI CCD to perform a fine resolution chirp transform.

bandwidth was limited to 10 KHz and the data presented in Fig. 7 is therefore only a portion of a full folded spectrum. Crosssections along the bias ridge are shown in Figs. 7a-d with different fine frequency inputs. The bias pedestal drops away near the edge of the output trace indicating the region of processor analysis $H=VT$. At a fine frequency difference of 120 Hz the constructive peak is quite large and well away from the edge. At a frequency of 240 Hz the destructive peak has moved over by about its width indicating that a resolution of 120 Hz or better has been achieved. At a frequency of 600 Hz the peak has moved over much farther but seems to be broadening. At 2100 Hz it is clear that the peak is broadening and decreasing in size. This appears to be due to the analog nature of the reference waveform generation introducing statistical frequency variation that produce larger and larger phase uncertainties at higher frequencies. This problem can be alleviated through the use of a digital buffer memory and digital to analog converter in order to produce the reference waveform. However this digital buffer would need to contain 1024 lines of 1024 words each that must be read out at a 100 MHz rate, which is currently unavailable to us. For this reason an alternative architecture with a much simpler reference function was built for comparison purposes.

Time Integrating Chirp Transform on TDI CCD

A common method of performing spectral analysis with analog convolvers is the chirp transform algorithm. This algorithm results from the observation that the Fourier kernel can be decomposed into a convolution and two products through the use of the following identity.

$$-2nm = (n-m)^2 - n^2 - m^2 \quad (17)$$

This expression can be substituted into Eq. 10, which defines the desired DFT, and rearranged to yield the chirp transform expression.

$$S(m) = \sum_{n=0}^{N-1} s(n) e^{-j2\pi nm/N} = \exp(-j\pi m^2/N) \sum_{n=0}^{N-1} [s(n) \exp(-j\pi n^2/N)] \exp(j\pi (n-m)^2/N) \quad (18)$$

An examination of this representation indicates that we can accomplish a DFT by first pre-multiplying the input signal with a chirp, convolving the resulting product with a matched chirp, and finally post multiplying the convolver output with a matched chirp in the Fourier domain. Often only the power spectrum is required in which case we can neglect the final post multiplication by the quadratic phase factor, since its magnitude is unity everywhere.

There are a large number of methods for performing convolutions with time integrating optical techniques utilizing acousto-optics and charge coupled devices. The features which must be incorporated in a convolver are an array of multiplications, a shift operation and an integration. An architecture we have used with great success for chirp correlations¹⁰ requires a temporally modulated input, a fixed chirp reference mask, and a CCD operated in the time delay and integrate (TDI) mode, where the detector time integrates in a sliding coordinate frame. The CCD readout circuitry is modified so that it operates in a continuously scrolling fashion referred to as TDI or shift and add. In this manner the CCD accumulates photogenerated charge at each pixel site for one or more laser pulses, the charge is then transferred vertically (y) by one pixel of width Δy , and each pixel accumulates more charge, adding it to that previously detected. By the time a pixel has fully traversed the array of height Y , it has accumulated photogenerated charge from each position in space at successively delayed times. If a temporally modulated sampled signal $f(nT)$ is uniformly incident on the chirp mask $c(y) = \cos(\pi by^2) \text{rect}(y/Y)$, and then imaged on the TDI CCD the resulting signal that scrolls off the the edge of the detector is given by the convolution of the temporal signal with the spatially sampled image of the mask.

$$I(mT) = \sum_n f(nT) c(m\Delta y) * \delta(m-n) = \sum_n f(nT) c((m-n)\Delta y) \quad (19)$$

The index m enumerates the output samples obtained at the edge of the CCD at successive times mT . We can use this convolver to configure a time integrating interferometric chirp transform power spectrum analyzer by introducing a quadratically phase modulated reference beam, and interfering this with a baseband phase modulated signal beam in order

to perform an interferometric chirp pre multiplication of the input signal as the detected intensity $f(nT) = s(nT) \cos(b(nT)V)$, where $V = \Delta y/T$ is the TDI CCD shifting velocity. The resulting time integrated power spectrum output would be obtained by demodulating the interferometric product term from the bias terms and squaring the chirp transform electronically. The analysis bandwidth of this type of TDI chirp transform spectrum analyzer is limited by the sampled mask resolution to half the CCD shift rate for each sideband of the reference chirp. A two sided chirp reference can cover an analysis bandwidth equal to the CCD shift rate. A feature of this type of chirp transform is that the spectral amplitude decreases linearly away from the center frequency as does the spectral resolution, due to a decrease in the matching chirp overlap.

The chirp transform spectrum analyzer can be readily incorporated in the TSI folded spectrum processor by including a quadratically phase modulated reference beam, and a chirp reference mask imaged onto a TDI CCD. The resulting system architecture is shown schematically in Fig. 8 and can be seen to be just a modification of the previous optical system. The analytical description of the TSI folded spectrum system with the TDI chirp transform fine resolution processors results in the same signal and bias terms as in the previous system and shown in Fig. 6. However the TDI approach has the advantageous feature that it averages coherent optical artifacts and CCD pixel noise along the shifting column thereby minimizing the system fixed pattern noise. The system output is given by

$$F(u, m) = \sum_{n=0}^{N-1} \left| \int_A f(x, n) e^{-j2\pi ux} dx + r \exp[-jb(nT)^2 V^2] \right|^2 \cos(b(m\Delta y)^2) \text{rect}[m\Delta y/Y] \cdot \delta(m-n) \\ + \text{bias terms} \quad (20)$$

$$= \sum_{n=0}^{N-1} \left[\int_A f(x, n) e^{-j2\pi ux} dx \right] 2r \cos(b(n\Delta y)^2) \cos(b((m-n)\Delta y)^2) \text{rect}[(m-n)\Delta y/Y]$$

Chirp Transform Output

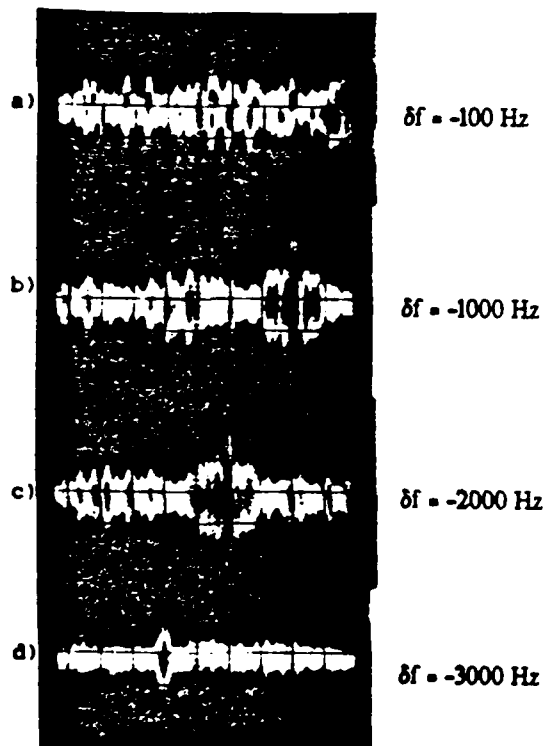


Fig. 9. Crosssections through coarse frequency ridge for the folded spectrum processor of Fig. 8 for fine frequencies a) 100 Hz b) 1000 Hz c) 2000 Hz d) 3000 Hz.

2-D Output

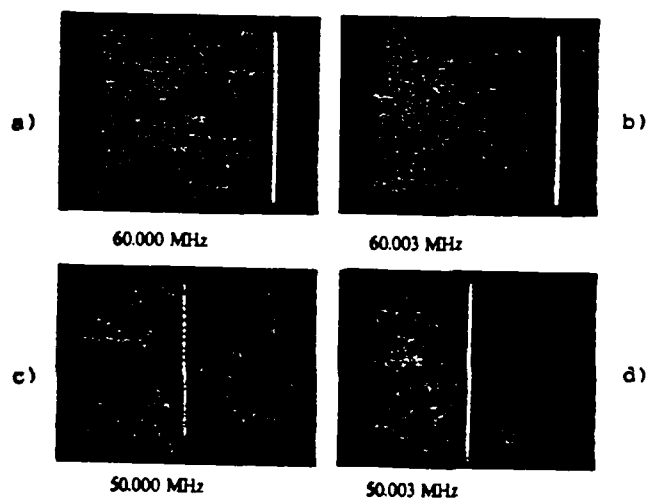


Fig. 10. Impulse response of TSI folded spectrum processor displayed in a 2-D format on a video monitor. a) 60.000 MHz b) 60.003 MHz c) 50.000 MHz d) 50.003 MHz.

We can see that the quadratically phase modulated reference beam results in an interferometric premultiplication of each sampled baseband coarse frequency bin with a temporal chirp. The combination of the chirp mask and the TDI CCD results in a chirp convolution for each coarse frequency channel, and the resulting folded spectrum raster scans out of the CCD as it scrolls. Experimental results from this processor are shown in Fig. 9 and Fig. 10. Figures 9a-d show crosssections through the coarse frequency ridge of the impulse response of Fig. 6 for several input fine frequencies, and the decrease of peak height with increasing frequency is apparent. Figure 10 shows the 2-D Folded Spectrum output displayed on a video monitor for a variety of input frequencies, where the coarse ridge is seen to move with 10 MHz frequency change, and 3 KHz moves the fringes in the fine dimension along the ridge. The spatial and temporal chirp references used in these experiments were two sided with 9 fringes per side giving a compression of 36, and resulting in a spectrum analysis with 36 independently resolvable frequencies. The sideband structure is highly visible due to the hard clipping of the mask, both the spectral resolution and the linearity are better than 100 Hz, and the performance closely matches the expected theoretical behavior.

Demodulation of Signal From Bias Terms

A crucial requirement for the successful operation of a TSI signal processing system is the ability to demodulate the signal from the bias terms. There are two main approaches to signal demodulation which are shown in Fig. 11, spatial carrier demodulation and CCD based demodulation. In carrier demodulation the interferometric signal term is placed on a spatial carrier by the adjustment of the reference beam incidence angle, while the bias terms remain at baseband. When these signals are output from the CCD, a bandpass filter can separate the different components, and allow measurement of spectral amplitude and phase. The filtered output can be electronically squared and low passed in order to obtain a power spectrum, or rectification followed by a low pass can be used for measuring spectral amplitude components. The CCD raster readout mechanism can be used to change a spatial carrier in the horizontal direction into a fast temporal carrier that can be simply bandpassed by an active filter. A spatial carrier in the vertical dimension requires a more complicated set of video delay lines and a tapped filter in order to implement the desired bandpass operation. The spatial carrier can be placed in the coarse or fine frequency directions, but will cut down the available CCD space bandwidth product in whichever dimension it is used. In order to avoid losing detector resolution, another possible approach is to have two synchronously operating CCD detectors, one to detect the interferometric signal and bias terms, and the other with just the signal term in order to detect the signal dependent bias. The two detectors must be precisely aligned and subtracted pixel by pixel in order to remove the signal dependent bias, the reference beam bias term should be uniform across the array and can be subtracted with a level shifter. An experimental bias subtraction technique we are investigating requires a special purpose CCD with a Fill and Spill structure abutted to the transfer register in order to bleed off an appropriate amount of charge to leave the signal term and a small amount of bias, rather than have the bias use up the available CCD dynamic range.

AO Folded Spectrum Processor Performance

A summary of the performance parameters for the TSI folded spectrum processor described in this paper are given in Fig. 12, for the Caltech prototype, and for the potential of a fully engineered state of the art brassboard. The processor analysis bandwidth is determined by the AOD bandwidth, the CCD width and Fourier lens focal length, and by the laser diode pulse width at an acceptable level of illumination and temporal coherence. The Caltech prototype was limited by the CCD width, which can easily be modified. The minimum laser diode pulse width is determined by the multimode transient operation for the first 1 nsec which results in incoherent operation inappropriate for this processor, however light limitations will also dictate a minimum pulse width determined by peak output power. The number of coarse frequency bins is determined primarily by the number of CCD pixels, but another limitation is due to the AOD time bandwidth product and apodization uniformity. The availability of the Tektronix 2000x2000 CCD is very encouraging for the development of optical processors requiring high resolution 2-D detectors. The coarse frequency resolution is limited by the AOD transit time and apodization, but also determined by the FT lens focal length and pixel width. The number of fine frequency resolution bins is determined by the space bandwidth of the reference functions and the number of CCD pixels available. If a spatial carrier is included in the fine frequency dimension, then the number of fine resolution bins is further limited, and a maximum spatial carrier frequency of 4-6 pixels per cycle indicates that a maximum of 500 fine frequency bins could be predicted. The fine resolution is fundamentally limited by the CCD integration time during which the LD is being pulsed and the prototype processor has closely approached this limit. Actually a greater difficulty is encountered in attempting to achieve a fine frequency analysis

bandwidth equal to the coarse frequency resolution, which requires high bandwidth electronic reference generation currently not obtainable with our prototype circuitry. The TDI chirp transform approach is limited in analysis bandwidth by the mask resolution and by the obtainable CCD shift rate, which is currently limited to 15 KHz for video compatible devices, but could be increased with special CCD circuitry. The total number of independent spectral measurements that are being made in our initial experiments was only 15,000 but a potential for on the order of 10^6 clearly exists. The dynamic range is currently quite small due to coherence limitations and light starvation, but this will be greatly improved in the next generation processor. The most important performance parameter for the SETI search is the sensitivity to a signal buried in noise, and the processing gain limitations of the analog technique used to perform the spectral analysis. The large AOD dynamic range allows very weak signals to be effectively transduced into the optical processor, but a huge CCD dynamic range is not required to achieve the desired processing gain, since the spatial integration achieves much of it.

Since sensitivity and processing gain are so important to the SETI application we will include a simple analysis of the processor limitations based on the model presented in Fig. 13. For a weak narrowband CW input buried in white noise with power σ_n we will have a system input

$$s(t) = a e^{j\omega t} + n(t). \quad (21)$$

The amplifier and AOD have a large linear region and a saturation regime that will introduce third order intermodulation products that will degrade system performance. An ideal gain G is included due to an optical power increase which will not increase the relative power in the intermodulation terms. The processor consists of a SI coarse frequency channelizer which acts as 10^3 band pass filters (BPF), cascaded onto an array of 10^3 TI spectrum analyzers, resulting in the synthesis of 10^6 narrowband filters. The detected spectral amplitudes are corrupted by the inclusion of signal dependent noise, and the detector noise term with variance σ_d . The processing gain is defined as the ratio of input signal to noise ratio (SNR) to output SNR, and ideally it would be given by

$$PG = \frac{SNR_{in}}{SNR_{out}} = \frac{a^2/\sigma_n^2}{a^2/\beta^2\sigma_n^2} = 1/\beta^2 \sim 10^6. \quad (22)$$

Where β^2 is the BPF width divided by the entire bandwidth, which is equivalently the inverse of the number of channels. After the demodulation operation, which will remove the bias terms and including the bias shot noise contribution in σ_d , the expected value of the power spectrum component at the frequency of the input tone will be

$$\langle S_{\omega}^2 \rangle = a^2 G^2 + \beta^2 G^2 \sigma_n^2 + \sigma_d^2. \quad (23)$$

The input sensitivity is determined by the minimum detectable signal at the desired false alarm probability, which for SETI is on the order of 10^{-5} to 10^{-6} , in order to keep the total number of false alarms manageable. By cross scanning the sky and comparing positions and frequencies of spectral components exceeding the given threshold, the false alarm probability can easily be squared. Thus the minimum detectable signal resulting in a false alarm probability of e^{-2} would be 10 standard deviations above the noise.

$$a_{min}^2 G^2 > 10[\beta^2 G^2 \sigma_n^2 + \sigma_d^2] \quad (24)$$

To minimize effect of the detector noise limitations we increase the gain G to the point that the signal dependent noise and the detector noise are comparable. In this case minimum input SNR that can result in a signal exceeding the threshold is given by

$$SNR_{min} = a_{min}^2/\sigma_n^2 = 10[\beta^2 + \sigma_d^2/G^2\sigma_n^2] \sim 20\beta^2. \quad (25)$$

Thus the resulting input sensitivity is degraded by 13dB from the initial -60dB, but only 3dB of the degradation is an artifact of the analog optical processor. This estimate for the actual input sensitivity of -47dB will be further degraded by the intermodulation product terms due to the amplifier and AOD nonlinearities, by the shot noise due to the bias, and probably most importantly by light starvation limiting the gain G .

Conclusion

In this paper we have presented preliminary experimental results from time and space integrating folded spectrum processors that may have potential application in the SETI program. The sampling action of the repetitively pulsed laser diode eliminates the need for a distributed local oscillator reference wave by aliasing all of the fine frequency components to baseband at the output of the coarse frequency channelizing space integrating spectrum analyzer as is required for the following CCD based narrow band time integrating spectrum analyzers. Two methods of time integrating fine frequency analysis were presented, the time integrating DFT, and the time delay and integrate chirp transform. The TI DFT has the attractive possibility for incoherent addition of spectra in order to improve SNR, but requires a very complex and accurate reference waveform. The TDI chirp transform helps average out coherent noise and detector noise, and has a simple reference function, but does not allow simple complex spectral demodulation, and is bandwidth limited by the available CCD shift rate. The possible advantages of this technique over alternative digital approaches, are a decrease in system cost and complexity, along with the elimination of the high speed analog to digital converter required for digital processing. For a possible future spaceborne SETI search the optical processor has decreased size, weight and power requirements over that of a digital approach.

This work was performed under grants from the AFOSR and NASA. Kelvin Wagner is the recipient of an ARO graduate research fellowship at Caltech. We thank Sam Gulkis and the entire SETI science working group for introducing us to SETI and partially motivating this work. We also thank Mike Haney for his collaboration on TSI optical processing.

REFERENCES

- [1] F. Drake, J. E. Wolfe, and C. L. Seeger, Eds., 'SETI Science Working Group Report,' NASA Tech Paper 2244, Oct. (1983).
- [2] C. E. Thomas, 'Optical Spectrum Analysis of Large Space Bandwidth Signals,' Applied Optics, vol. 5, p. 1782, (1966).
- [3] T. M. Turpin, 'Spectrum Analysis Using Optical Processing,' Proc. IEEE, vol. 69, p. 79, Jan. (1981).
- [4] P. Kellman, 'Time Integrating Optical Signal Processing,' Ph.D. dissertation, Stanford Univ., Stanford Ca., (1979).
- [5] T. R. Bader, 'Coherent Optical Hybrid Techniques For Spectrum Analysis,' Proc. SPIE, vol. 185, p. 140, (1979).
- [6] T. R. Bader, 'Acousto-optic Spectrum Analysis: A High Performance Hybrid Technique,' Applied Optics, vol. 18, p. 1668, (1979).
- [7] D. Psaltis and D. Casasent, 'Time and Space Integrating Spectrum Analyzer,' Applied Optics, vol. 18, p. 3203, (1979).
- [8] D. Psaltis, 'Two Dimensional Optical Processing Using One-Dimensional Input Devices,' Proc. IEEE, vol. 72, p. 962, July (1984).
- [9] A. Van der Lugt, 'Interferometric Spectrum Analyzer,' Applied Optics, vol. 20, p. 2770, (1981).
- [10] D. Psaltis and K. Wagner, 'Real-Time Optical SAR Processor,' Optical Eng., vol. 21, p. 822, (1982).
- [11] M. Haney and D. Psaltis 'Measurement of the Temporal Coherence Properties of Pulsed Laser Diodes,' Applied Optics, vol. 24, p. 1926, (1985).

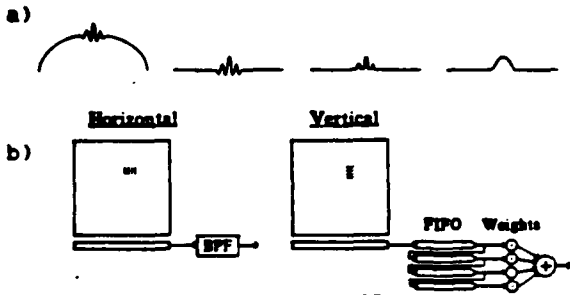
AO PROCESSOR PERFORMANCE

Demodulation Schemes

Required to separate signal from bias terms

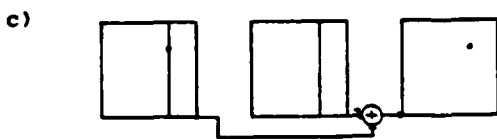
Carrier Demodulation

Signal + bias Bandpass Filter Rectification Envelope Detection



Bias Subtraction with 2 CCDs

signal + bias bias signal



Fill and Spill adaptive bias subtraction CCD

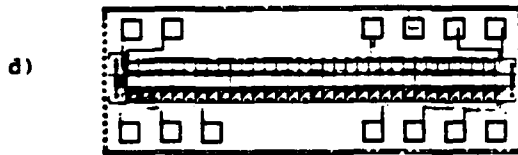


Fig. 11 a) Carrier demodulation. b) Implementations for horizontal or vertical carriers. c) Bias subtraction with 2 CCDs. d) Fill and spill bias subtraction CCD test chip.

	Caltech Prototype	Potential
B = Bandwidth	30 MHz	1 GHz
n_c = # Coarse Bins	384	2000
f_c = Coarse Resolution	80 KHz	B/1000
n_f = # Fine Bins	491/ a_{carrier} 36	2000/ a_{carrier}
f_f = Fine Resolution	1 KHz 120Hz	$f_c/1000$
N = Spectral Resolution	15,000	10^6
DR = Dynamic Range	50 levels	1000 levels
Sensitivity	-40dB	-60dB

Fig. 12. TSI AO processor performance for Caltech prototype and predicted state of the art brassboard processor.

Sensitivity and Processing Gain

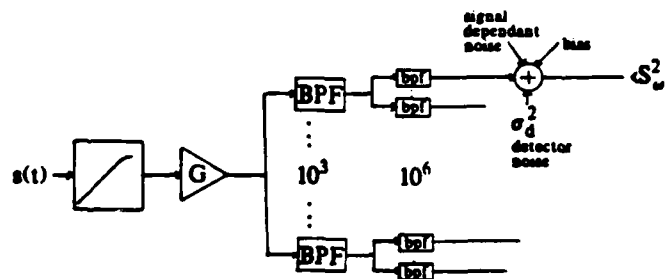


Fig. 13. Model used for sensitivity and processing gain analysis.

Real Time Computation of Moments with Acousto-optics

Kelvin Wagner and Demetri Psaltis

Department of Electrical Engineering
 California Institute of Technology
 Pasadena, CA 91125

Abstract

The separable nature of the geometric moment generating functions and the serial raster encoding of the image permits the 2-dimensional moment integral equation to be computed with a cascaded system of 1-dimensional integrations. In this paper new optical processing architectures are presented that use acousto-optic devices illuminated by pulsed sources to enter the data into the optical system. The advanced state of the art of the components used and the flexibility of these architectures can lead to the implementation of practical optical processing system for computing the moments of a real time high resolution image.

Introduction

Given a finite image intensity distribution $f(x,y)$ over an aperture A , the p,q th geometric moment m_{pq} of order $p+q$ is given by the moment integral equation

$$m_{pq} = \iint_A x^p y^q f(x,y) dx dy. \quad (1)$$

The value of these geometric moments are sensitive to the position, scale, rotation and contrast of the image. Appropriate nonlinear combinations of moments can be made invariant to all of these intraclass variations.^{1,2,3,5,6,8} These moment invariants may have applications in pattern recognition as a low dimensionality feature space for image classification⁴ and in coding the essential features of an image⁸. Alternatively the moments can be used to normalize an image⁷ so that intraclass variations are minimized, so that standard pattern recognition techniques such as template matching can be applied successfully. The zeroth order moment gives the average of $f(x,y)$. The first order moments locate the image centroid $\bar{x} = m_{10}/m_{00}$, $\bar{y} = m_{01}/m_{00}$. The central moments μ_{pq} are defined with the centroid as the origin and are invariant to translations of the image

$$\begin{aligned} \mu_{pq} &= \iint_A (x-\bar{x})^p (y-\bar{y})^q f(x,y) dx dy \\ &= \sum_{r=0}^p \sum_{s=0}^q \binom{p}{r} \binom{q}{s} (-\bar{x})^{p-r} (-\bar{y})^{q-s} m_{rs}. \end{aligned} \quad (2)$$

Rotation and scale invariant moments of order (p,q) can be defined in terms of the ordinary moments m_{rs} .

Coherent optical processors have been proposed to compute the moments⁹⁻¹², using 2-dimensional spatial light modulators (SLM) for real time operation. The separable nature of the geometric moment generating functions and serial raster encoding of the image allows the decomposition of the moment integral into cascaded 1-dimensional integrations. New architectures are presented that use 1-dimensional acousto-optic (AOD) spatial light modulators to enter the data into the optical system. In the first architecture the separable moment generating functions are entered through the AOD's, which allows flexibility in the choice of generating functions, and accuracy of their representation. Alternatively the moment generating functions can be written on film or used to modulate the optical source and the raster recorded image is entered into the optical system through the AOD. The large dynamic range and rapid programmability of acousto-optic devices can allow the implementation of practical real time optical computers for the generation of the geometric moments of an image.

Space Integrating Processor

Film based coherent optical processors have been proposed⁹⁻¹² for computing the geometric moments of an image using the 2-D spatial integration performed by a spherical Fourier transform (FT) lens. Teague's⁹ method requires multiple differentiation of the FT of the image. The other methods^{10,12} involve the spatial integration of the product of the

image $f(x,y)$ and the moment generating function $g_{pq}(x,y) = x^p y^q$. Spatial frequency multiplexing of the generating functions in a computer generated hologram can be used to compute all the desired moments in parallel at separate locations in the frequency plane.

Since the geometric moment generating function $x^p y^q$ is separable in cartesian coordinates, we can replace the moment mask of Ref. 10,11 with a pair of crossed acousto-optic deflectors (AOD) schlieren imaged onto the image film as shown in Fig. 1. This is similar to a triple product processor architecture¹⁷, except that the continuously modulated source is now pulsed. If the pulse width is less than the inverse bandwidth of the signals in the AODs, then it will act as a delta function in time and freeze the image of the moment generating function when it illuminates the input image. The functions t^p and t^q are synchronously applied to the AODs as amplitude modulations to an RF carrier at the center frequency of the AOD

$$\begin{aligned} S_{1p}(t) &= t^p e^{j^2 \pi f_0 t} \\ S_{2q}(t) &= t^q e^{j^2 \pi f_0 t} \end{aligned} \quad (3)$$

For odd powers there is a change of sign at $t=0$ that is represented with a 180° phase change. This will lead to destructive interference at the integrated output which can be interferometrically detected. When these signals simultaneously slide to the centers of the AODs the laser diode (LD) is pulsed, collimated, and focused into a horizontal strip that enters AOD1 at the Bragg angle. The diffracted light is collimated in y and Fourier transformed in x to produce a vertical strip that illuminates AOD2 at its Bragg angle. Both undiffracted components are filtered, and the doubly diffracted light is expanded in x to a square so that AOD1 is imaged onto $f(x,y)$ in the $-x$ direction and AOD2 is imaged onto $f(x,y)$ in the $-y$ direction. The light amplitude incident on the image is similar to an outer product and is modulated by

$$\begin{aligned} E_i(x,y,t) &= \delta(t) S_{1p}(x+vt) S_{2q}(y+vt) \\ &= \delta(t) S_{1p}(x) S_{2q}(y). \end{aligned} \quad (4)$$

This is multiplied by the image transmittivity, $f(x,y)$, and the product is Fourier transformed by the integrating lens. This transform can be sampled at its center or integrated over its whole extent to yield an interferometrically detected signal proportional to the desired moment.

$$m_{pq} = \iint_A S_{1p}(x) S_{2q}(y) f(x,y) dx dy = \iint_A x^p y^q f(x,y) dx dy. \quad (5)$$

All the desired moments could be computed successively with high accuracy and dynamic range on a single detector pair very rapidly. For an AOD with a 10 μ sec aperture time moments through the tenth order could be computed in .5msec. This would allow many differently windowed sets of moments to be computed during each frame time of available 2-D spatial light modulators. Alternatively the moments could be computed simultaneously with lower dynamic range in a few microseconds by using temporal frequency multiplexing of the moment generating functions. This is shown in Figure 2. In this case the following signals are applied to the two AOD's of the system.

$$\begin{aligned} S_1(t) &= \sum_{p=0}^P t^p e^{j2\pi(f_0 + p\Delta f)t} \\ S_2(t) &= \sum_{q=0}^Q t^q e^{j2\pi(f_0 + q\Delta f)t}, \end{aligned} \quad (6)$$

where Δf is the frequency separation of the frequency multiplexed moment generating functions. For an image bandwidth B [cy/mm] and acoustic velocity v , the different moments would be spatially separated in the output plane if $\Delta f/v > B$, and therefore could be sampled by an appropriately arranged array of detectors.

The versatility of this system allows the direct optical computation of the central moments μ_{pq} instead of the geometric moments m_{pq} . First the centroid is located in each dimension separately by finding the 0th and 1st order moments and using $\bar{x} = m_{10}/m_{00}$.

$\bar{y} = m_{01}/m_{00}$, or by operating the source CW, and locating the zero crossing of the detector's temporal output as the generating functions for m_{10} or m_{01} slide through the Bragg cells.

To compute the central moments directly the moment generating functions are input to AOD1 and AOD2 with a relative delay $\tau_d = \tau_1 - \tau_2 = x/v - y/v$ and the LD is pulsed at $t = \tau_1$. This aligns the origin of the moment generating functions with the image centroid, and the spatially integrated outputs compute the central moments μ_{pq} . This architecture also allows the windowing of the region for which the moments are being computed by simply windowing the electronic signals $S_1(t)$ and $S_2(t)$, this would allow a rapid search for the appropriate window within $f(x,y)$ to be made under computer control.

Space and Time Integrating Architecture

The performance of the space integrating moment processor could be limited by the 2-D spatial light modulator (SLM) required for real time operation. In many applications the image is available to the optical processor in the form of a serial raster signal. In this case, the 2-D image $f(x,y)$ can be entered into the optical processor one line at a time through a 1-D SLM such as an acousto-optic device (AOD) ^{13,14}. The raster scanned representation of a continuous image $f(x,y)$ is denoted by $f(x, n\Delta y)$ where n is an integer and Δy is the separation between the raster lines. The moments can be calculated from the sampled image by replacing the integration over y in Eq. (1) with a summation over n

$$m_{pq} = \sum_{n=1}^N (n\Delta y)^q \int x^p f(x, n\Delta y) dx. \quad (7)$$

This formation suggest a hybrid space and time integrating system ^{16,15,13} as shown in Fig. 3. The video raster lines are sequentially applied to an AOD operated in the intensity modulation regime and oriented in the x direction, as an amplitude modulation of an RF carrier at the center frequency of the piezoelectric transducer. The AOD is illuminated by the collimated light emitted from a closely spaced array of $Q+1$ pulsed modulated laser diodes (LD). If all the LD's emit light at the same wavelength then the spatial frequency multiplexed light incident on the AOD must be within its Bragg angular aperture. Alternatively, the momentum mismatch can be compensated by varying the wavelength in the LD array, so that the light from each LD is incident on the AOD at the Bragg angle for that wavelength. When the n th video raster line completely fills the AOD aperture, the entire LD array is synchronously pulsed with a pulse width short enough to freeze the image of the raster line within the AOD. The intensity modulation of the q th LD is given by

$$I_q(t) = \sum_{n=1}^N (n^q + \text{bias}) \delta(t - nT), \quad (8)$$

where T is the video line time and the bias may be needed to represent the bipolar odd moment generating functions. The bias is not necessary for the even moments and spatial multiplexing can be used rather than bias to represent the bipolar odd moments. The light diffracted by the AOD is expanded and collimated in y and schlieren imaged in x , so the intensity incident on P_3 from the q th LD during the n th raster line is

$$O_3(x,t) = I_q(t) f(t - \frac{x}{v}, n\Delta y) = n^q \delta(t - nT) f(\frac{x}{v}, n\Delta y). \quad (9)$$

At plane P_3 a mask is placed containing the intensity transmittance variations x^p at different y coordinates:

$$t(x,y) = \sum_{p=0}^P x^p \text{rect} \left[\frac{y - py_0}{y_0} \right]. \quad (10)$$

The light leaving P_3 (modulated by the product of the incident intensity and the mask transmittance), is then imaged in the y dimension and integrated in the x direction onto the plane P_4 . At the output plane there are $(Q+1)$ columns of $(P+1)$ detectors each. Each LD, is imaged onto a separate column and to avoid crosstalk the columns must be separated by $\Delta > \lambda Fv/B$, where F is the focal length of the integrating lens, v is the acoustic velocity and B is the video bandwidth. If the detector is wide enough to collect all the light from its associated laser diode then the detected signal on the p,q th detector during the n th pulse will be proportional to

$$s_{pq}(n) = \int_{(n-\frac{1}{2})T}^{(n+\frac{1}{2})T} I_q(t) \int 0_3(x,t) x^p dx dt = n^q \int f\left(\frac{x}{V}, n\Delta y\right) x^p dx. \quad (11)$$

If the detector is time integrated for a full video frame the output of the p,qth detector becomes

$$\begin{aligned} m_{pq} &= \int_0^{NT} I_q(t) \int 0_3(x,t) x^p dx dt \\ &= \sum_{n=0}^N n^q \int f\left(\frac{x}{V}, n\Delta y\right) x^p dx \\ &= \sum_{n=0}^N \int n^q x^p f\left(\frac{x}{V}, n\Delta y\right) dx, \end{aligned} \quad (12)$$

which is proportional to the desired moment calculation. All $(P+1) \times (Q+1)$ moments are computed in parallel on a 2-dimensional array of detectors.

An alternative topology of the optical processor is obtained by spatially multiplexing the LDs in the y direction as shown in Fig. 4. This eliminates the problem of Bragg angle mismatch previously encountered, and allows all the moments to be detected on a single linear detector array, without crosstalk. The mask must be repeated once for each laser diode for this architecture and its intensity transmittance is given by

$$t(x,y) = \sum_{q=0}^Q \sum_{p=0}^P x^p \text{rect} \left[\frac{y - q(p+1)y_0 - py_0}{y_0} \right] \quad (13)$$

Another useful feature of this architecture is its extension to bipolar processing as shown in Fig. 5. For odd q two laser diodes can be spatially multiplexed in the x direction, one represents the positive and the other the negative portions of the moment generating function. Similarly for odd p two rows of the mask are used to represent the positive and negative portions of this half of the moment generating function. Two linear time integrating detector arrays are used to detect the four cross products. The outputs are electronically added and subtracted to yield the calculated bipolar moment.

Discussion

In optical analog computers of this type errors and noise can enter into the calculation for many reasons. These include noise in the modulated laser diode sources, nonlinearities in the acousto-optic Bragg cell, quantization errors and film nonlinearities of the computer generated masks, optical aberrations, coherent noise and detector noise. Careful optical and electronic design can minimize most of these effects, and in practice the output detector array is a major factor that limits the accuracy of an optical computer. Typical charge coupled device (CCD) detector arrays have an output SNR of 1000:1, which allows a digitization of the optically computed geometric moments to a resolution of 10 bits. A digital post processor can then combine the geometrical moments to produce the invariants and perform a pattern classification based on the invariant feature space. The geometrical moment optical preprocessor takes the bulk of the computational load away from the digital computer, however some accuracy is sacrificed.

The separable moment generating function μ^P is positive definite for even powers p, and an incoherent unipolar intensity representation can be used. For odd powers p, the moment generating functions become bipolar. Therefore the capability to represent bipolar quantities must be incorporated in the optical processor. If coherent light is used the cancellation needed for odd moment powers can be performed in the space integrating processors by destructive interference and interferometric detection. However in the time integrating architectures the processing is performed by integrating photogenerated charge on the detector, which is proportional to the number of incident photons or the intensity and therefore strictly positive. Bipolar operations can be performed by placing the odd moment generating functions on a bias. This in turn places a bias on the detector thereby decreasing the available dynamic range of the computed moments. Alternatively the positive

and negative halves of the separable odd moment generating functions can be represented in separate channels of the processor. The crossproduct terms must be electronically combined with the appropriate sign when the detector is read out in order to produce an estimate of the bipolar moment. The achievable accuracy and dynamic range of this bipolar estimate will be limited by the requirement that none of the detectors saturate, and that the variances of the noise will add. The possibility of a special purpose directly coupled detector structures to perform the continuous subtraction needed for incoherent bipolar optical processing merits further research.

Conclusions

In conclusion we have proposed several optical computing architectures for calculating the geometrical moments of images in real time. The space integrating processor requires a 2-D SLM to represent the image, but has the flexibility of electronic control of the separable moment generating function. With this processor the moments can be calculated sequentially at a rapid rate with high accuracy. The time and space integrating architectures are not as flexible since the moment generating functions are fixed by the computer generated masks. However since acousto-optic devices are used to enter the raster video data into the optical processing system, the need for a real time 2-D SLM is avoided.

Acknowledgments

We would like to thank Dr. Yaser Abu-Mostafa for many illumination discussions on moment invariants.

The research reported in this paper is supported by the Air Force Office of Scientific Research and the Army Research Office.

References

1. M. K. Hu, IRE Trans. Inf. Th. IT-8, 179 (1962).
2. S. Maitra, Proc IEEE 67, 697 (1979).
3. S. S. Reddi, IEEE Trans. Pattern Anal. Machine Intell. PAMI-3, 240 (1981).
4. S. Dudani et.al., IEEE Trans. Computers C-26, 39 (1977).
5. R. C. Gonzales and P. Wintz, Digital Image Processing, Addison-Wesley, Reading, Mass., (1977).
6. Y. Abu-Mostafa and D. Psaltis, IEEE Trans. on PAMI, to be published.
7. Y. Abu-Mostafa and D. Psaltis, IEEE Trans. on PAMI, to be published.
8. M. R. Teague, JOSA 70, 920 (1980).
9. M. R. Teague, Applied Optics 19, 1353 (1980).
10. J. A. Blodgett et.al., Optics Letters 7, 7 (1982).
11. D. Casasent and D. Psaltis, SPIE 201, 107 (1978).
12. D. Casasent and D. Psaltis, Optics Letters, 5, 395 (1980).
13. D. Psaltis, JOSA, 71, 198 (1981).
14. D. Psaltis, Applied Optics, 21, 491 (1982).
15. D. Psaltis and D. Casasent, Applied Optics, 18, 3203 (1979).
16. T. R. Bader, Applied Optics, 18, 1668 (1979).
17. P. Kellman, Ph.D. thesis, Stanford U. Press, Stanford, California (1979).

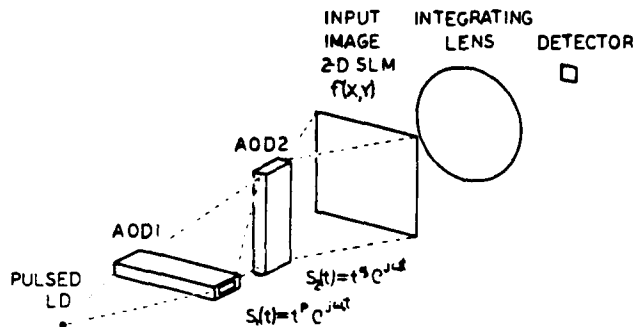


Figure 1. Space integrating geometric moment processor with crossed Bragg cells to enter the separable moment generating function.

Space Integrating Moments Architecture with Crossed Bragg Cells

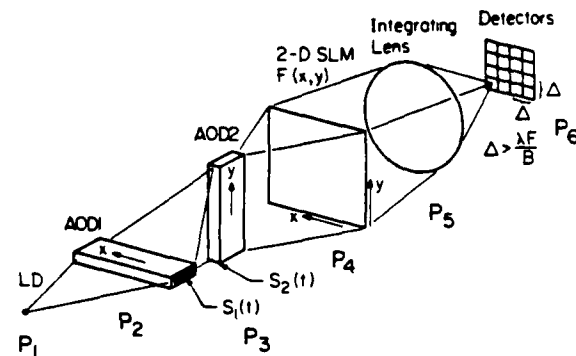


Figure 2. Frequency multiplexed space integrating architecture.

Space and Time Integrating Moments Architecture

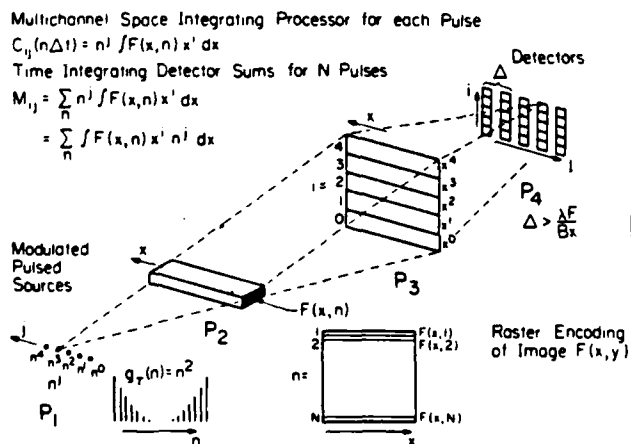


Figure 3a. Time and space integrating moment processor with horizontal spatial multiplexing of modulated pulsed laser diodes and acousto-optic entry of video raster data.

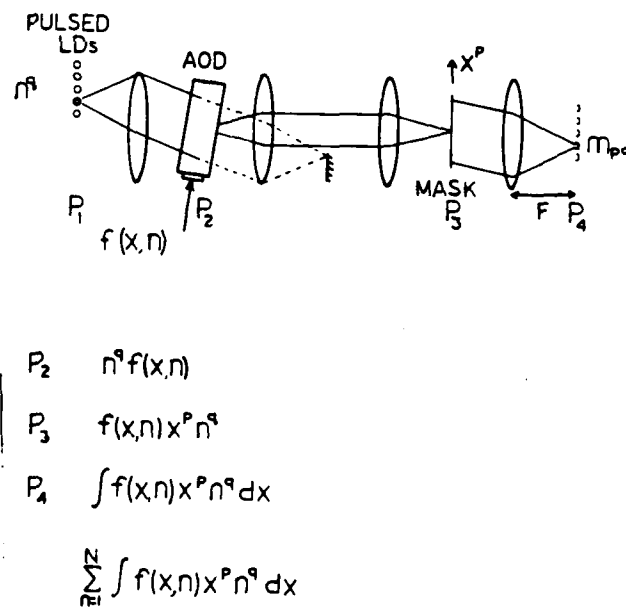


Figure 3b. Top view and operation of the processor.

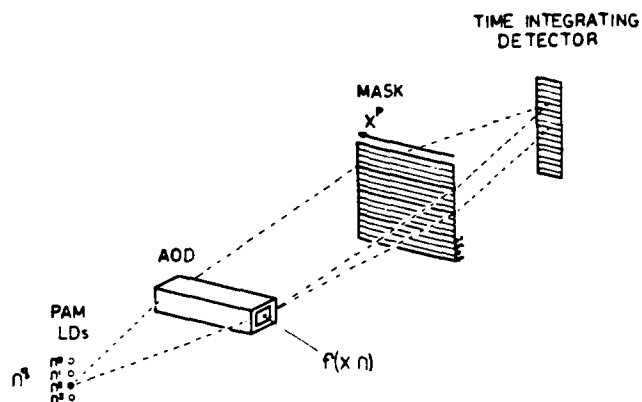


Figure 4. Time and space integrating moment processor with vertical spatial multiplexing of the laser diode array.

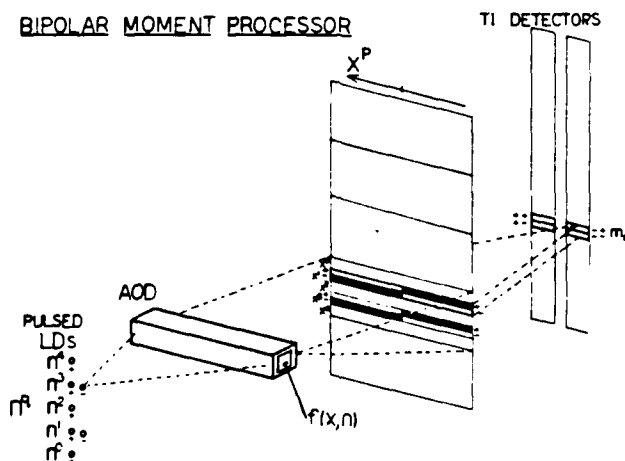


Figure 5. Bipolar representation of the odd moment generating function in the processor of Figure 4, by horizontal multiplexing of the sources and vertical multiplexing in the mask.

A SPACE INTEGRATING ACOUSTO-OPTIC MATRIX-MATRIX MULTIPLIER

Kelvin WAGNER and Demetri PSALTIS

*Department of Electrical Engineering, California Institute of Technology,
Pasadena, CA 91125, USA*

Received 9 August 1984

An optical architecture is described for performing pipelined matrix-matrix multiplications. The architecture is implemented using multiple transducer acousto-optic devices and a wideband photodetector array. A variant of engagement formatting allows multiple inner products to be simultaneously computed by 1-D spatial integration, and through proper pipelining the full product matrix is produced at the output of the detector array. The output matrix in this architecture is in a format that is directly compatible with the input, a feature that can facilitate the implementation of iterative matrix algorithms. Digital multiplication by analog convolution can be incorporated for improved accuracy by using a frequency multiplexed representation of the binary data.

1. Introduction

The optical implementation of matrix operations has received considerable attention recently. Architectures and algorithms have been designed that have increased the speed, accuracy and flexibility of optical matrix processors, extending the potential applicability of such systems to a broader range of problems. Specific advances that have been accomplished in recent years include the initial demonstration of vector-matrix multiplication [1], the introduction of time-integrating systolic [2], engagement [3,4], and outer product optical processors [5], a frequency multiplexed processor [6], improvements in accuracy with residue arithmetic [7], the utilization of the method of digital multiplication by analog convolution (DMAC) [8,9] in the above array processors [10,11], and a combination of systolic processing and the DMAC algorithm in a two dimensional implementation utilizing crossed multichannel Bragg cells for matrix vector multiplications with digital accuracy [12].

Perhaps the most significant application of numerical optical processors is in $O(N^3)$ problems, i.e. matrix algebra problems that require a minimum of N^3 multiplications and additions, where N is the size of the matrix involved. The solution of a set of linear equations, matrix inversion, and singular value decom-

position are examples of such problems [13]. Optical techniques can be applied to such problems by selecting an algorithm that can implement the required operation with successive matrix-matrix multiplications, such as Gauss eliminations, Givens transformations or Householder reflections [13]. For these algorithms N optical matrix multiplications are required, and since each matrix-matrix multiplication requires N^3 multiplies and adds, optical systems usually solve an N^3 problem with N^4 operations. However, the speed and parallelism of optics can make the optical implementation advantageous, despite this inefficiency. The product matrix that is produced at each iteration during the execution of such an algorithm is used as one of the input matrices for the next iteration. It is therefore important that the format of the output product matrix is directly compatible with the input in order to avoid reformatting and minimize the iteration time. The architecture described in this paper was selected principally because the output can be amplified and applied directly to the input. A space integrating implementation using a parallel output wideband photodetector array is chosen for accuracy and speed considerations. Several candidate data flow optical architectures satisfying these requirements are possible. In this paper we present one such data flow matrix processor which uses crossed multichannel acousto-

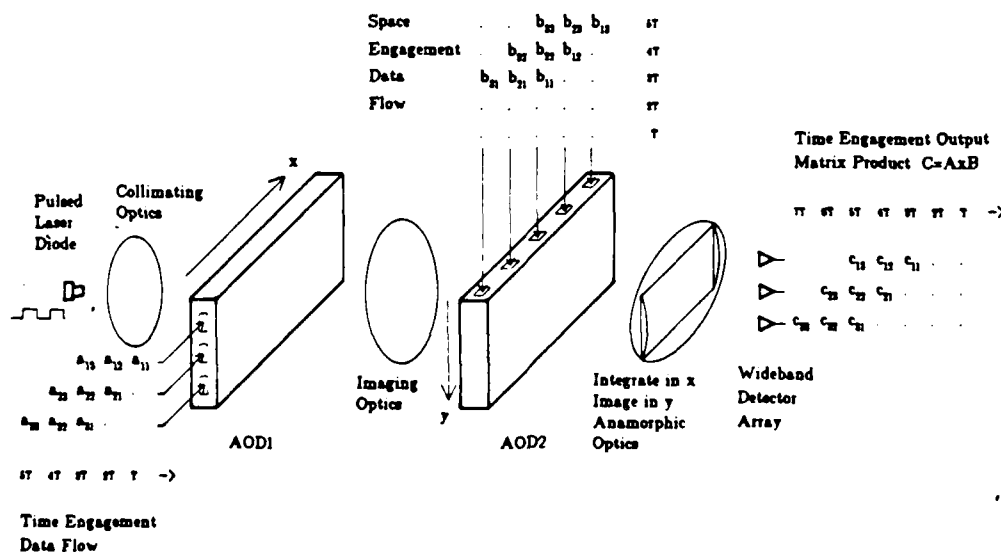
optic devices (AOD). The operation of the system is based on time and space alignment of vectors which allows the formation of multiple inner product summations via spatial integration in a pipelined fashion. The matrix format is similar to an engagement array, but the data flow is transposed so that the local multiplications needed for each inner product operation form in parallel in space, rather than sequentially in time. Each inner product is summed by a 1-D space integrating condensing cylinder onto an output detector. Many such inner product accumulators are multiplexed in the orthogonal dimension onto separate detectors of a linear array.

2. Optical processor architecture

The proposed optical architecture for matrix-matrix multiplication is shown in fig. 1, along with the appropriate data flow. The principal components of the system are a pulsed laser, two orthogonal multichannel Bragg cells (one with N channels, and the other with $2N - 1$ channels), a linear array of N wideband photodetectors, and lenses. The optical processor is a 2-D array of N^2 analog multipliers configured as an array of N space integrating inner product processors. With the appropriate engagement format of the ma-

trices an array of inner products is formed on the detector array during each processor cycle. As data flows through the Bragg cells the output appears in the same engagement format as the input, which allows direct feedback for iterative operations without latency.

Global system synchronization and sample definition at the output are provided at each time interval T , by the strobing action of a repetitively pulsed laser diode. The pulsed light is collimated and incident on the first multichannel acousto-optic device, AOD1, at the Bragg angle in the x dimension. The elements of an $N \times N$ matrix A are applied to the N transducers of AOD1 in an engagement representation. They propagate continuously along the x direction at a velocity equal to one inter transducer spacing of AOD2 each T s. Rows are represented in individual channels as sequential acoustic pulses separated by T s. The n th row of the matrix A is applied to the n th transducer of AOD1 with a delay of nT s. In this manner the matrix is folded back in time into a sliding parallelogram format we call time engagement. The optical field emerging from AOD1 is spatially filtered in the Fourier plane to remove the undiffracted component and the diffracted field is imaged onto AOD2 at the Bragg angle in y . For clarity the image reversal of the imaging system is ignored. Matrix B propagates in AOD2 in the y direction one channel separation of AOD1



each T s, orthogonally to, and synchronously with, the motion of A . B is delayed by $(N - 1)T$ s with respect to A to allow the first row of A to fully enter AOD1. At time $(N - 1 + k)T$, the k th column of the matrix B is applied to transducers k through $k + N$ of AOD2. In this representation, called space engagement, the matrix is folded over in space to $2N - 1$ parallel channels which require N time cycles to be completed. The doubly diffracted light is imaged in y and space integrated in x by the anamorphic lens system following AOD2. The light collected on each photodetector during each cycle is the sum of the product of the elements that are aligned within the corresponding horizontal channels of the two AODs. As we will see in the following paragraph, at the output of the detector array we obtain the product matrix $C = A \times B$ in a time engagement format.

At the N th time increment the first row of matrix A in AOD1 and the first column of matrix B in AOD2 are spatially aligned in the top channel of the system. The N local products $a_{1i}b_{i1}$ are calculated in parallel by imaging AOD1 through AOD2, and the sum $c_{11} = \sum_{i=1}^N a_{1i}b_{i1}$ is produced by spatially integrating all these products onto the top detector. One time increment T later the second row of A has fully entered the second channel of AOD1, and simultaneously the first row of A has moved one column away from the transducer. In the orthogonal dimension, the first column of B has moved down in AOD2 one channel to engage the second row of A and produce $c_{21} = \sum_{i=1}^N a_{2i}b_{i1}$ via space integration onto the second detector. Concurrently the second column of B has been applied to transducers 2 through $N + 1$, in order to align with the first row of A and produce $c_{12} = \sum_{i=1}^N a_{1i}b_{i2}$ on the first detector. In a similar manner successive inner products are aligned, locally multiplied and globally accumulated to the N output detectors. In a total of $2N - 1$ time increments T , the output product matrix is produced in a time engagement format identical to the format of the matrix A . Therefore it can be fed directly back to the N transducers of AOD1 with no reformatting or latency. This allows for fully efficient pipelining of iterative matrix algorithms, since after the matrix A is initially loaded into AOD1 there are no more waiting periods required to load new matrices. For instance, when the first element of the output matrix c_{11} is produced the first row of A has been fully entered into AOD1 so we can begin entering the first row of the output matrix in the top channel of

AOD1. No interference with subsequent calculations will occur because of the zeroes included in the space engagement formatting of matrix B .

If the Bragg cells are operated in the linear amplitude diffraction regime, and coherent detection is used, then it is possible to make the outputs of the photodetectors appear at the center frequency of AOD1, simplifying direct feedback. The coherent implementation allows complex valued matrices to be represented by the magnitude and phase of the acoustic pulses. If the Bragg cells are operated in the linear intensity diffraction regime (incoherent implementation), then only real positive matrices can be represented, but simpler non-interferometric detection can be employed. In this case the output matrix would not be on a carrier, therefore mixers would be required to upconvert the output before amplifying and applying to AOD1.

3. Frequency multiplexed DMAC

An increase in the accuracy of this processor can be incorporated by the use of the digital multiplication by analog convolution algorithm (DMAC) [8-12], at the expense of additional complexity. It is well known that the multiplication of two time domain waveforms results in the convolution of their Fourier spectra. This can be utilized to implement the DMAC algorithm by simply multiplying the frequency multiplexed binary representations of two numbers.

The product $z = x * y$ of two M bit integers x and y can be expressed as follows:

$$\begin{aligned} z &= \sum_{k=0}^{2(M-1)} z_k 2^k = \left(\sum_{i=0}^{M-1} x_i 2^i \right) \left(\sum_{j=0}^{M-1} y_j 2^j \right) \\ &= \sum_{k=0}^{2(M-1)} 2^k \left(\sum_{i=0}^{M-1} x_i y_{k-i} \right), \end{aligned} \quad (1)$$

where x_i, y_i are the bits in the binary representation of the integers x and y and the coefficients $z_k = \sum_{i=0}^{M-1} x_i y_{k-i}$ can achieve M discrete levels. A time domain representation of a frequency multiplexed binary word is given by $f(t) = \sum_{n=0}^{M-1} x_n \exp(jn\omega t)$. The product of two such waveforms is

$$h(t) = f(t)g(t)$$

AD-A173 411

ACOUSTO-OPTIC PROCESSING OF 2-D SIGNALS USING TEMPORAL
AND SPATIAL INTEGR. (U) CALIFORNIA INST OF TECH
PASADENA DEPT OF ELECTRICAL ENGINEER. D PSALTIS

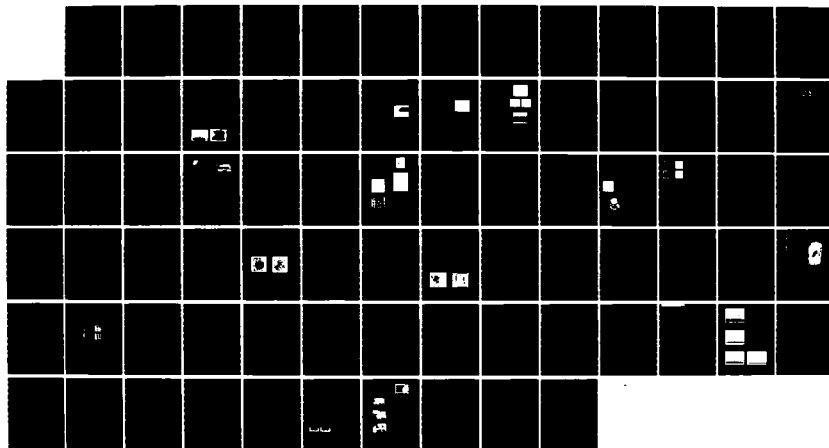
2/2

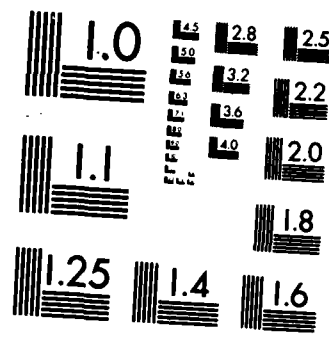
UNCLASSIFIED

29 APR 86 AFOSR-TR-86-0047 AFOSR-82-0128

F/G 17/9

ML





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

$$\begin{aligned}
&= \left(\sum_{n=0}^{M-1} x_n \exp(jn\omega t) \right) \left(\sum_{m=0}^{M-1} y_m \exp(jm\omega t) \right) \\
&= \sum_{k=0}^{2(M-1)} \left(\sum_{n=0}^{M-1} x_n y_{k-n} \right) \exp(jk\omega t) \\
&= \sum_{k=0}^{2(M-1)} z_k \exp(jk\omega t). \quad (2)
\end{aligned}$$

Thus the weights of the frequency multiplexed product waveform correspond to an M level digitally weighted representation of the product of the two binary words. This pseudo binary representation can be channelized into $2M - 1$ frequency channels, each centered at $k\omega$. The amplitude of each spectral component can be quantized to one of M levels by an A/D converter, and true binary representation can be obtained with a digital shift and add register.

Digital multiplication by frequency convolution can be incorporated in the matrix multiplier of fig. 1 in order to improve the accuracy over that attainable with analog data representation. This will increase the required time bandwidth product of the AODs by a factor of at least the number of bits. The duration of the optical and acoustic pulses must be at least $2\pi/\omega$ s, to permit channelization at the detector output. The frequency multiplexed binary weighted data must be encoded in phase within each acoustic pulse so that all the frequency components add constructively. When interferometric detection is used, the RF output from each detector will be the coherent sum of N frequency multiplexed multilevel binary weighted signals, occupying up to twice the original bandwidth. This format is compatible with direct feedback to AOD1 without redigitization in each cycle. Further iterations would increase the required dynamic range of each frequency component and increase the number of nonzero frequencies. By examining the Fourier plane of AOD1 we can determine globally the number of frequencies occupied for matrix A . If the available number of frequency bins is exceeded by p excess Fourier components, then we can perform a global pseudo floating point rescaling of the product matrix by increasing the local oscillator frequency used for heterodyne detection by $p\omega$. The lower order p bits are then discarded by highpass filtering the detector outputs, or with a Fourier plane aperture. Redigitization is required if a detectors dynamic range is ex-

ceeded or when the iterative matrix operation has converged, and the resultant matrix must be output in binary form.

The major difficulty with incorporating DMAC in the matrix-matrix multiplier is the complexity of the frequency multiplexed encoding and decoding, and the large number of A/D converters required to convert back to true binary form. Each of the N detectors must be channelized into $2M$ frequency bands, and each of these must be digitized every T s to an accuracy of NM . This could require as many as $2NM$ A/Ds, which is probably impractical. Alternatively, N faster A/Ds could be multiplexed between $2M$ frequency channels each by performing a conversion each $T/2M$ s. To decrease the number of A/Ds even further it may be possible to digitize only one detector output at a time, and recycle the matrix for further conversions by multiplying by the identity matrix I .

4. Discussion

Acousto-optic devices are attractive transducers for data flow optical processors because of their sliding window nature, wide bandwidth ($>1\text{GHz}$), large number of resolvable spots ($TB > 1000$) and wide dynamic range ($>60\text{dB}$). When dealing with multichannel Bragg cells, however, the constraints imposed by acoustic diffraction and electrical crosstalk limit all aspects of device performance. Today a practical limit on the number of channels is on the order of 100 or less. Eventually larger arrays may be realized through the use of anisotropic self collimation and effective RF isolation techniques. The properties of the multichannel acousto-optic devices determine the processing power of the matrix multiplier and to a lesser degree the accuracy obtainable. In order to perform an $N \times N$ analog matrix multiplication this architecture would require an N channel AOD with $TB = 2(2N - 1)$ (AOD1), and a $2N - 1$ channel AOD with $TB = 2N$ (AOD2), where a factor of 2 was included for intra pulse dead time. If frequency multiplexed binary encoding is used with M bits, then AOD1 requires a $TB > M(2N - 1)$, and AOD2 requires a $TB > NM$.

The detector array is composed of N parallel wide-band photo-detectors whose outputs are bandpass filtered, combined with a steering matrix, amplified, and fed back to AOD1. Single photodetectors can have a wide dynamic range ($>50\text{dB}$), but large monolithically

fabricated arrays are limited by crosstalk. To obtain the full dynamic range capabilities of the detectors, a powerful optical source must be employed, and optical losses minimized.

To obtain an estimate of the processing power of the space integrating optical matrix multiplier described in this letter, let us consider a target system for multiplying 32×32 matrices. AOD1 requires 32 channels and a TB > 128, OAD2 requires 63 channels and a TB > 64. If we assume a bandwidth of 128 MHz, then the full matrix product could be obtained in 1 μ s, yielding an analog processing rate of 3.2×10^{10} multiplications per second. At these rates the matrix could be inverted in as little as 32 μ s with a direct algorithm, or a fraction of a millisecond with an iterative algorithm. If we desire digital accuracy of 8 bits input and 16 bits output, then the system parameters become much more stringent. AOD1 requires a TB > 1024, and AOD2 requires a TB > 512, which could be accomplished with a bandwidth of 100 MHz, and a matrix multiplication time of approximately 10 μ s. This yields a processing rate of 3×10^9 DMAC multiplications per second. However, to obtain this additional accuracy we require an array of 32 frequency channelizers with 16 frequency bins each, and 32 8-bit A/D operating at 100 MHz, temporally multiplexed between the frequency channels. When the complexity of the electronic peripherals to the optical processor reaches this high level, it is important to keep in perspective what would be required for a fully digital implementation. For instance, for the same numbers as above a digital array of 32 8-bit multipliers/accumulators each operating at 100 MHz can have the same processing power (3.2×10^9 operations/s). In other words, comparable electronic hardware is needed for both the digital and the DMAC optical implementations. The advantage of the optical implementation is in the reduced communication rate (320×10^6 samples/s versus 3.2×10^9 samples/s in this example) between the high speed array processor and the system buffer memory. This advantage derives from the 2-D parallelism of optics which results in 1024 parallel multiplications being performed each 100 ns interval.

5. Conclusion

A highly parallel, pipelined, space integrating, acousto-optic processor for iterative matrix-matrix

multiplication has been described. The architecture avoids the serial readout bottleneck and dynamic range limitations of CCD arrays used in time integrating architectures. The wideband nature of the input and output transducers can result in an analog processing rate exceeding 30 GOPs (Billion multiplies per second), and in excess of 3 GOPs for the DMAC implementation. Additional accuracy can be incorporated by the use of the DMAC algorithm and frequency multiplexing, but the improved accuracy is accompanied by an increase in complexity and expense. The analog optical processor can be implemented with currently available devices and simple electronic support circuitry. It provides extremely high processing power with reasonably good accuracy (equivalent to 8–10 bits) due to the high dynamic range that is achievable with non integrating photodiode arrays.

Acknowledgements

The research reported here is supported by the Air Force Office of Scientific Research and the Rome Air Development Center. K. Wagner is the recipient of an Army Research Office graduate fellowship. Illuminating discussions with T. Weverka and E. Miles are gratefully acknowledged.

References

- [1] J.W. Goodman, A.R. Dias and L.M. Woody, *Optics Lett.* 2 (1978) 1.
- [2] H.J. Caulfield, W.T. Rhodes, M.J. Foster and S. Horvitz, *Optics Comm.* 40 (1982) 86.
- [3] P.S. Guilfoyle, *Proc. SPIE* 352 (1982) 2.
- [4] R.P. Bocker, *Appl. Optics* 22 (1983) 804.
- [5] R.A. Athale and W.C. Collins, *Appl. Optics* 21 (1982) 2089.
- [6] D. Casasent, J. Jackson and C. Neuman, *Appl. Optics* 22 (1983) 115.
- [7] J. Jackson and D. Casasent, *Appl. Optics* 22 (1983) 2817.
- [8] H.J. Whitehouse and J.M. Speiser, in: *Aspects of signal processing*, pt 2, ed. G. Tacconi (Proc. NATO Advanced Study Institute) Boston (1976) pp. 669–702.
- [9] D. Psaltis, D. Casasent, D. Neff and M. Carlotto, *Proc. SPIE* 232 (1980) 151.
- [10] R.P. Bocker, S.R. Clayton and K. Bromley, *Appl. Optics* 22 (1983) 2019.
- [11] R.A. Athale, W.C. Collins and P.D. Stilwell, *Appl. Optics* 22 (1983) 368.
- [12] P.S. Guilfoyle, *Opt. Eng.* 23 (1984) 20.
- [13] G.H. Golub and C.F. VanLoan, *Matrix computations* (John Hopkins University Press, Baltimore, 1983).

Adaptive Acoustooptic Processor

Demetri Psaltis and John Hong
Department of Electrical Engineering, California Institute of Technology
Pasadena, Calif. 91125

Abstract

Space-integrating, acoustooptic processors for adaptive, temporal filtering are examined. The basic architecture is then extended to the space-time domain for application in broadband phased array processing. An acoustooptic processor capable of such 2-dimensional, adaptive processing is described.

I. Introduction

A major portion of optimum filtering theory concerns itself with the efficient separation of useful signals from additive noise. Fixed optimum filters, such as the Wiener filter, are applicable when the signal and noise statistics are stationary and known apriori; the lack of such apriori knowledge, however, motivates the implementation of adaptive filters which estimate the required signal and noise characteristics. The implementation of such filtering techniques requires a processor which must be able to compute various correlation functions of signal and noise and change its transfer function accordingly.

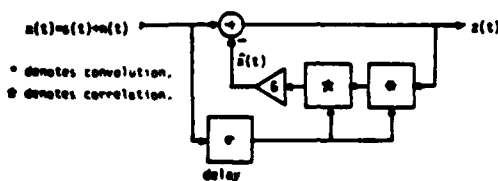
The situation becomes considerably more complex when one is required to adaptively filter signals in the space-time domain, as in the case of broadband array processing. The number of broadband jammers that an adaptive processor can cancel without compromising its directional discrimination can be used as its performance measure. Electronic implementations of such array processors¹ have exhibited limited performance with respect to this measure. The transversal filters responsible for temporally filtering the outputs from the array elements are of low order and hence, operation is usually limited to narrowband signals, due to hardware limitations.

The bandwidth requirements and the parallel nature of array processing make optical implementations attractive. Various optical implementations have been explored by researchers in the area^{2,3}. In a recent paper⁴, we described two optical adaptive filters for use in the time domain. These implementations are strictly one-dimensional, requiring only one dimensional devices, permitting an extension to the space-time domain through the use of multi-channel AOD's. In this paper, we briefly review the operation of the temporally adaptive acoustooptic filter. After this foundation has been established, an adaptive array processor utilizing multi-channel AOD's in a space integrating architecture will be described.

II. Adaptive Temporal Filters

To be adaptive, a processor that optimizes either the mean-square error or SNR criterion must compute various correlation functions of the input signal and noise and vary its filter response characteristics accordingly. Specifically, we will consider situations where the signal is broadband and the additive noise consists of strong, narrowband jammers whose frequencies are unknown; the jammer frequencies must be estimated for effective noise rejection.

Shown in Fig.1 is a system diagram of the Passive Processor which was shown to adaptively perform an approximate Wiener filtering operation⁴. The operation of the system can be explained heuristically in the following manner. Suppose that the input consists solely of a single sinusoid, and consider the feedback signal, $\hat{x}(t)$, with the loop opened at the summing junction. The input sinusoid is first convolved with a replica of itself, delayed by ϵ to produce another sinusoid of the same frequency, also delayed by ϵ , at the input of the correlator. This is then correlated with the input delayed by ϵ to produce a sinusoid with the same phase as the input sinusoid appearing at the summing junction; the delays, ϵ , cancel out due to the cascaded correlation-convolution operations. The possibility of stable cancellation is now apparent since the two signals, the input and feedback signals, which are subtracted at the junction, are identical in frequency and phase. Cancellation does not occur for a broadband input, because it correlates poorly with delayed versions of itself, resulting in a negligible feedback signal, $\hat{x}(t)$. The passive processor thus discriminates against narrowband components of the input signal, while preserving the desired, broadband components of the input, making it suitable for signal estimation.

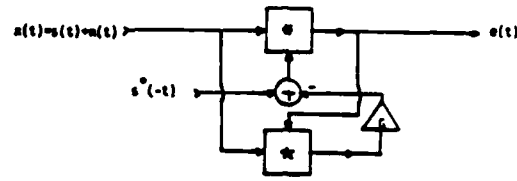


$$z(t) = x(t) - \int x(t-\tau) s^*(\tau) d\tau + s^*(t-t)$$

$$Z(\omega) = X(\omega) / (1 + |G(\omega)|^2)$$

approximates the Wiener Filter for low input SNR.

Fig.1 Passive Processor



$$e(t) = x(t) + [s^*(-t) - G(t)] \cdot s^*(-t)$$

$$E(\omega) = S^*(\omega)X(\omega) / (1 + |G(\omega)|^2)$$

approximates the maximum SNR Filter for low input SNR.

Fig.2 Active Processor

For signal detection, where a known broadband signal is to be detected amidst an additive collection of narrowband jammers of unknown frequencies, a different system is required. Shown in Fig. 2 is a system diagram of such a system, the Active Processor, along with a brief, mathematical description of its operation; a complete description can be found in reference 4. As with the Passive Processor, the basic operation of this system can be understood using heuristic arguments. Suppose that the input consists solely of a large amplitude sinusoid of frequency f_0 , and the desired waveform, $s(t)$, is a broadband signal. With the loop opened at the summing junction, the output, $e(t)$, will consist mostly of a sinusoid of frequency f_0 ; this signal is then correlated with the input to produce another sinusoid of the same frequency at the summing junction. When the loop is closed, the output sinusoid resulting from the convolution of the input and the code waveform, $s(t)$, is 180° out of phase with that which is produced by the feedback signal, and hence, cancellation of the sinusoid takes place. Again, the feedback signal is negligible for small amplitude broadband signals.

Both the Passive and Active systems require convolution and correlation as the basic building blocks. In choosing the architecture to be used, we consider the following important requirements: 1) wide bandwidth, 2) dynamically controllable convolver and correlator, 3) convolver - correlator compatibility, 4) large dynamic range. These considerations lead us to choose a space-integrating acoustooptic architecture.

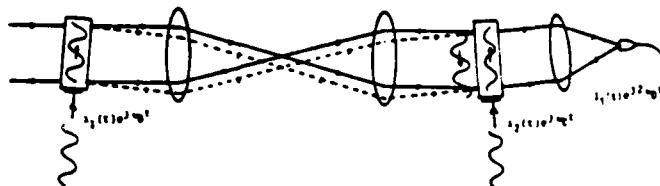


Fig.3 Space-Integrating Acoustooptic Convolver

Convolution can be performed with two AOD's as shown in Fig.3. The diffracted signal from the first AOD is imaged with unity magnification onto the second and, in the coherent realization shown, the diffracted components are spatially integrated onto a single detector whose output photocurrent is

$$I_1(t) = \int_{-T/2}^{T/2} i_1(t+\tau) i_2(t-\tau) d\tau, \quad \tau = x/v, \quad T = W/v. \quad (1)$$

$i_1(t)$ and $i_2(t)$ are the complex amplitudes of the AOD input signals, v is the acoustic velocity in the AOD, and W is the physical length of the AOD aperture. This integral can be manipulated to yield the more familiar form:

$$I_1(t) = \int_{t-T/2}^{t+T/2} i_1(\tau) i_2(2t-\tau) d\tau, \quad (2)$$

which is recognized to be a finite window convolution, time-compressed by a factor of two. The seemingly troublesome compression is actually an advantage in this case as will be apparent when we consider the correlator implementation.

Correlation can be performed with the convolution configuration just described by time-reversing the input to one of the AOD's. This, however, is not acceptable for real-time operation. Shown in Fig. 4 is a space-integrating processor which provides the relative motion between the two input signals that is required for correlation by using a demagnifying system between the two AOD's. The diffracted light from the first AOD is imaged with a 2:1 magnification onto the second AOD, and the diffracted light from both AOD's is spatially integrated onto a single detector. If the input to the second AOD is time-compressed by a factor of two, i.e., $i_4(t) = i_3(2t)$, then the correlator yields

$$I_2(t) = \int_{-T/4}^{T/4} i_3^*(t+2\tau) i_4(t+\tau) d\tau = \int_{t-T/2}^{t+T/2} i_3^*(\tau) i_5(t+\tau) d\tau, \quad (3)$$

as its output. The above is seen to be a finite window correlation of the two signals, $i_3(t)$ and $i_5(t)$. The advantage of obtaining the time compressed output from the convolver is now clear, since it can be used as the input to the correlator just described and thus obtain a consistent, cascaded convolution-correlation operation.

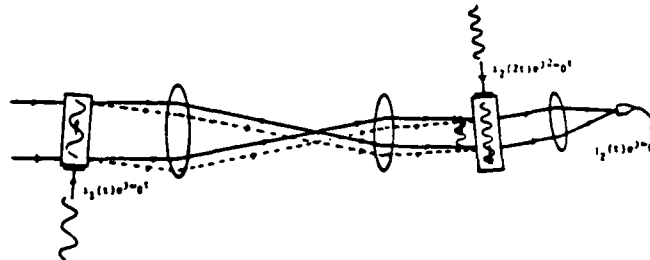


Fig. 4 Space-Integrating Acoustooptic Correlator

The optical implementations of the Passive and Active Processors are shown in Figs. 5 and 6, respectively, and they differ only in their electrical interconnections. Fig. 5(6) is the system of Fig. 1(2) with the optically implemented blocks of Figs. 3 and 4. In both cases, the upper branch of the processor computes the correlation while the lower performs the convolution. Since the convolver and correlator have a common input, the first AOD is shared. A more detailed description can be found in the previous publication⁴.

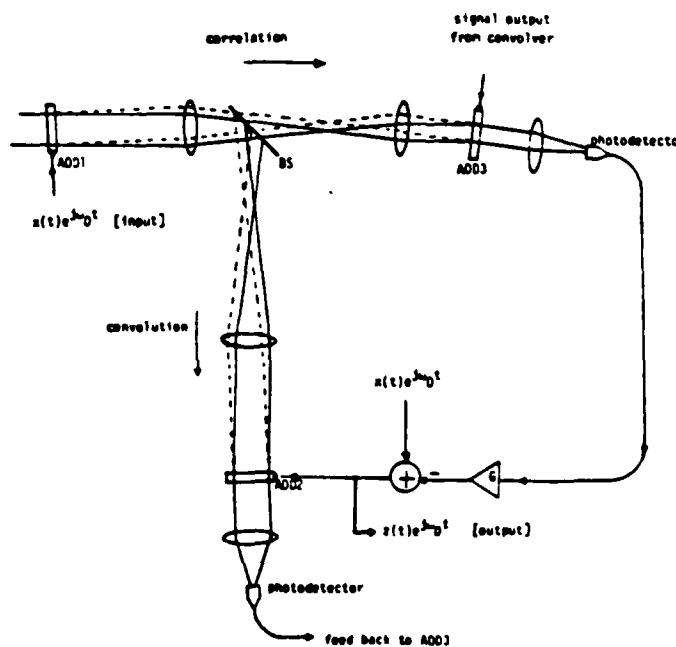


Fig. 5 Adaptive Optical Processor (Passive)

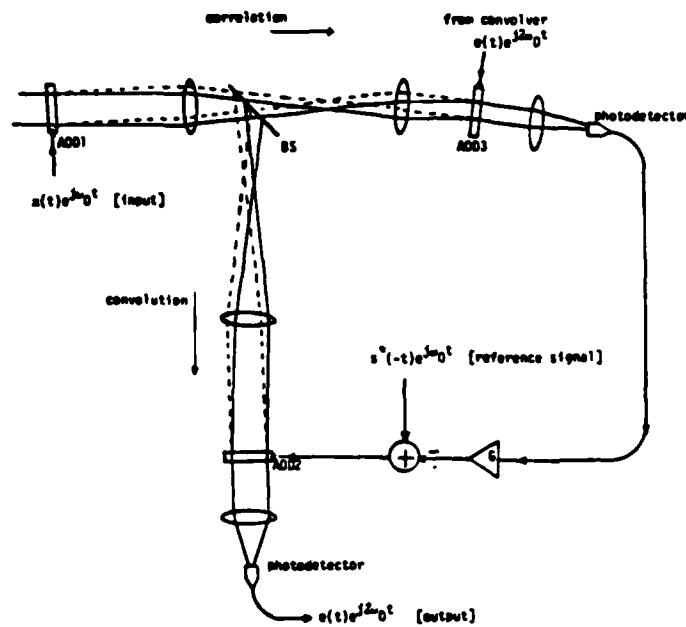


Fig.6 Adaptive Optical Processor (Active)

III. Adaptive Acoustooptic Array Processor

A general adaptive scheme is shown in Fig.7 in which each channel corresponding to one array element is filtered by a transversal filter with n controllable tap weights; if the array has N antenna elements, then the array processor requires the adaptive control of nN independent weights. Therefore, the computational load for broadband adaptive phased array processing is much heavier than that for adaptive temporal filtering because here, the array processor must adapt in two independent dimensions: space and time.

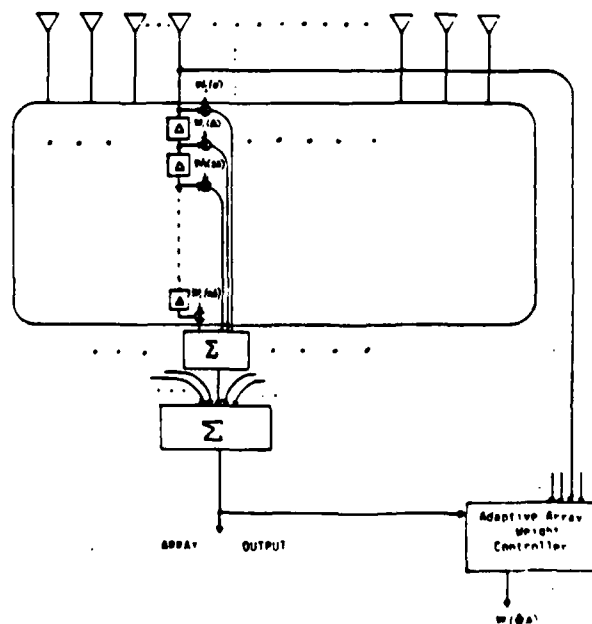


Fig.7 General Adaptive Array Processing Scheme

Optical processors are particularly attractive candidates for adaptive broadband array signal processing because of their inherent, multidimensional processing capability. Moreover, the broadband requirement can be met by the use of broadband multi-channel AOD's; the large time-bandwidth products that are available translate to the possibility of implementing very high order transversal filters. The acoustooptic processor which we now describe is an extension of the active temporal processor described earlier, to the space-time domain. Here, we employ a combination of multi and single channel AOD's to perform the required operations.

The output of a general, space-time filter, with a finite accumulation time can be expressed as

$$y(t) = \sum_{i=1}^N \int_{-T/2}^{T/2} x_i(\tau) h_i(t-\tau) d\tau, \quad (4)$$

where $x_i(t)$ is the signal from the i th antenna element, and $h_i(t)$ is the filtering function in each channel. A similar expression is obtained for the optically implemented space-time filter using two multi-channel AOD's shown in Fig.8. The output of this filter is

$$y(t) = \sum_{i=1}^N (1/T) \int_{-T/2}^{T/2} x_i(t-\tau) h_i(t+\tau) d\tau, \quad (5)$$

and we see that the only difference from the general filter (Eq.4) is in the time compression of the output. It can be shown that the optimum choice for $h_i(t)$ in Eq.(5) satisfies the system of integral equations:

$$\sum_{j=1}^N \int_{-T/2}^{T/2} h_j^*(t+\tau') r_{ij}(\tau-\tau') d\tau' = \lambda s_i(t-\tau), \quad (6)$$

where $s_i(t)$ is the desired signal vector and $r_{ij}(t)$ is the covariance matrix of the input noise given by

$$r_{ij}(\tau) = E [n_i(t) n_j^*(t-\tau)]. \quad (7)$$

$n_i(t)$ is the noise vector appearing at the array elements.

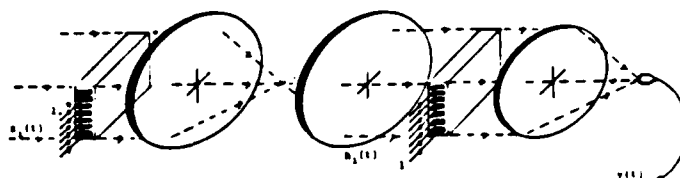


Fig.8 Acoustooptic Space-Time Filter

For adaptivity, we need to calculate and continuously update the filter function $h_i(t)$ to drive the output to the optimum result. As with the previously described Active Processor, the output must be correlated with the input to produce the appropriate filter function. Since the array processor has n inputs and one output, this requires that we correlate n signals with a common one. This can be achieved with the arrangement shown in Fig.9 which shows the use of a multi-channel AOD driven by the n antenna element outputs, in conjunction with a single channel AOD which is driven by the array output signal. Specifically, the output of the i th element of the linear detector array of the correlator is given by

$$r_i(t) = (1/T) \int_{-T/2}^{T/2} y(t+\tau) x_i^*(t+\tau) d\tau. \quad (8)$$

For proper correlation to appear at each output, the signal driving the single channel AOD must be time-compressed by a factor of two. This is indeed the case for the system described, and thus, the AOD implemented space-time filter and the n -channel correlator with a single reference are compatible.

Shown in Fig.10 is the array processor system diagram that shows the interconnections that are required; it is a direct extension of the Active Processor to 2-dimensions. The output from each antenna element is correlated with the processor output to produce the filter function for that element. The steering vector, $s_i(-t)$, determines the look

direction of the array and also is the temporal reference signal used for the detection of the desired signal, $s(t)$. It can be derived by feeding a tapped delay line with $s(t)$; the output from the i th tap corresponds to $s_i(t)$ and the tap spacing is adjusted to obtain the proper look direction:

$$s_i(t) = s(t - i\Delta) = s(t - i(d/c)\cos\epsilon), \quad (9)$$

where ϵ is the desired look angle from boresight, d is the array element spacing, and c is the speed of light.

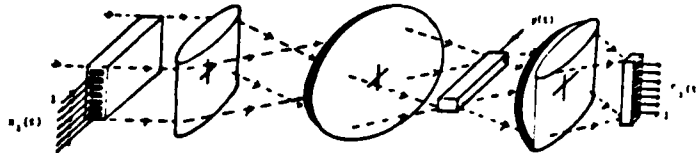


Fig. 9 Multi-Channel Acousto-optic Correlator

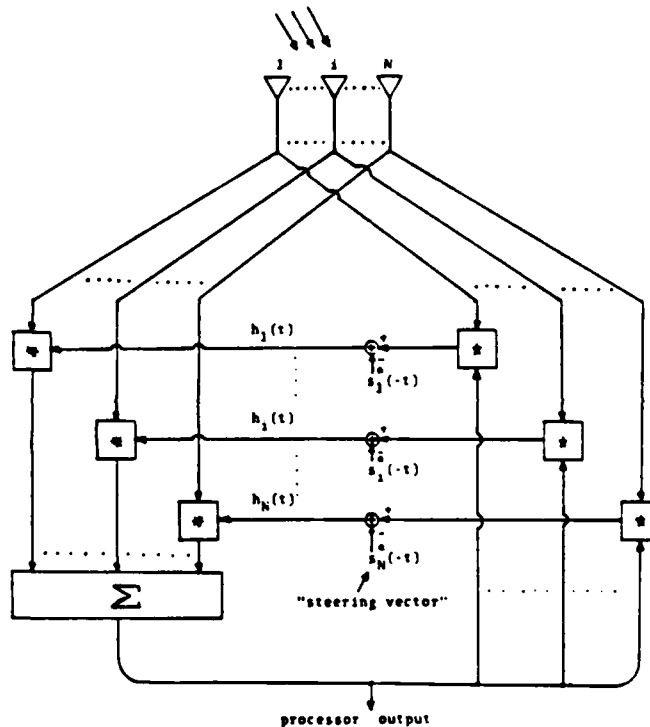


Fig. 10 Broadband Adaptive Array Processor

Fig. 11 shows the optically implemented adaptive array processor with the AOD-implemented space-time filter in the upper branch and the n -channel correlator in the lower one. By combining Eqs. (7) and (8), the equation that determines the filter function, $h_i(t)$ is seen to be

$$h_i(t) = s_i^*(-t) - (G/T^2) \sum_{j=1}^N \int_{-T/2}^{T/2} x_i^*(t+2\beta) x_j(t-\tau+\beta) h_j(t+\tau+\beta) d\beta d\tau, \quad (10)$$

where G is the feedback gain. Under conditions of low input SNR and large feedback gain⁴, Eq. (10) can be transformed to the frequency domain to yield

$$(G/4\pi^2) \sum_{j=1}^N H_j(\omega) \int_{-\infty}^{\infty} N_i^*(\omega_1) N_j(\omega_1) \text{sinc}^2[(\omega - \omega_1)T/(2\pi)] d\omega_1 = S_i^*(\omega), \quad (11)$$

For comparison, consider the Fourier Transform of Eq.6; when the accumulation time is infinite, the optimum filter equation is given by

$$2 \sum_{j=1}^N [\Gamma_{ij}(\omega) E_j(\omega) - \lambda^* S_1^*(\omega)], \quad (12)$$

$$[\Gamma_{ij}(\omega) = \text{F.T.} (\gamma_{ij}(\tau))],$$

where $[\Gamma_{ij}(\omega)]$ is the spectral density matrix. Identifying the integral in Eq.11 as the smoothed estimate of the spectral density matrix of the input noise vector, Eq. 11 is approximately equivalent to Eq.12. The effect of the finite time integration window is seen in the smoothing of the noise spectrum.

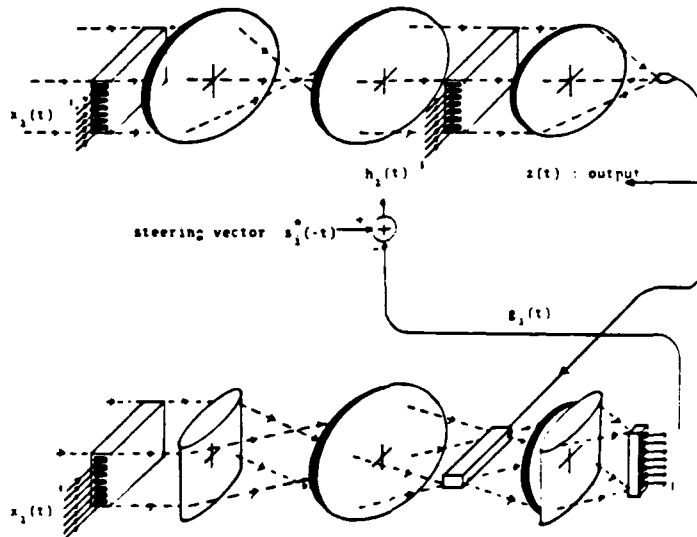


Fig.11 Adaptive Acoustooptic Phased Array Processor

Acknowledgement

This work is supported by the Rome Air Development Center and, in part, by the Air Force Office of Scientific Research.

IV. References

1. R. Riegler and R. Compton, Jr., 'An Adaptive Array for Interference Rejection,' Proc. IEEE 61, 748 (1973).
2. J. F. Rhodes, 'Adaptive Filter with a Time-Domain Implementation Using Correlation Cancellation Loops,' Appl. Opt. 22, 282 (1983).
3. A. Vanderlugt, 'Adaptive Optical Processor,' Appl. Opt. 21, 4005 (1982).
4. D. Psaltis and J. Hong, 'Adaptive Acoustooptic Filter,' Appl. Opt. 23, 3475 (1984).

Acoustooptic Adaptive Signal Processing

John Hong and Demetri Psaltis
California Institute of Technology, Department of Electrical Engineering
Pasadena, California 91125

Abstract

Acoustooptic methods for adaptive filtering of temporal signals are discussed. Two specific architectures are presented: one utilizing space-integration alone and the other combining both time and space integrating techniques. Performance issues regarding the space-integrating system are discussed in detail, and a description and experimental results for the space-time integrating filter are presented.

Background

Optimal filtering in unknown or time-varying signal environments requires the use of systems which can compute the necessary noise characteristics and adapt to yield approximately optimum performance. In this paper, the case of broadband signals received in the presence of strong, additive, narrowband jammers whose spectra are unknown is considered. This situation is illustrated in Fig.1, showing the useful broadband signal overwhelmed by a strong jammer. If the spectral density of the jammer were known apriori, then the Wiener filter whose transfer function is given by

$$H(\omega) = \frac{S_s(\omega)}{S_s(\omega) + S_n(\omega)} \quad (1)$$

would be used to yield the optimum performance in the least mean square error sense ($S_s(\omega)$ and $S_n(\omega)$ are the spectral densities of the signal and uncorrelated jammer noise, respectively). In the case of interest, the noise spectrum is unknown and must therefore be computed and the system response adjusted accordingly to approximate the Wiener filter.

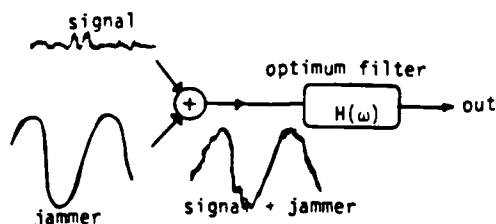


Fig.1 Optimum Filtering

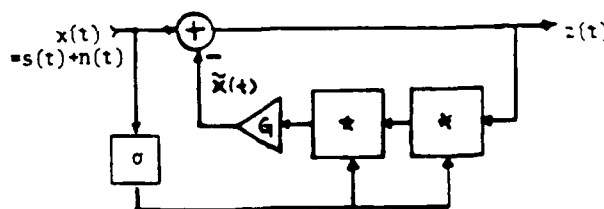


Fig.2 Passive Processor

Several optically implemented adaptive systems have previously been proposed^{2,5,6,8}. A block diagram of the adaptive system discussed in this paper is shown in Fig.2. The system has been described in detail in a previous publication², and its operation is briefly reviewed here: The two major components of the system are the convolution and correlation blocks (indicated by \cdot and \star , respectively, in Fig.2). The outputs of the convolver and correlator blocks are related to their inputs by

$$\begin{aligned} u(t) &= \int_{-\infty}^{\infty} s_1(t-\tau) s_2(\tau) d\tau, \\ w(t) &= \int_{-\infty}^{\infty} s_3(t+\tau) s_4^*(\tau) d\tau, \end{aligned} \quad (2)$$

respectively, where $s_i(t)$ denote the inputs to the blocks. The processor discriminates between narrowband signals which correlate well with delayed versions of themselves and broadband signals which correlate poorly. Thus, with the loop open at the summing node, the feedback signal $\tilde{x}(t)$ for an input consisting entirely of a broadband signal is negligibly small because such an input correlates poorly with delayed portions of itself. This is contrasted to the case where the input is a sinusoid, in which case the convolution of the input with itself delayed, followed by a correlation with the input delayed by the same amount yields a sinusoid of the same frequency and phase as the input. When the loop is closed, broadband components of the input feed through unaffected while suppression of correlated portions takes effect. The following approximate input-output relationship can be derived for the filter in Fig.2².

$$z(t) = x(t) - G \int_{-\infty}^{\infty} z(\tau) n^*(e-\epsilon) n(t+e-\tau-e) d\epsilon d\tau, \quad (3)$$

where $z(t)$ is the output, $n(t)$ is the input jammer noise, and $x(t) = s(t) + n(t)$ is the total input. A simple interpretation is found by the Fourier transformation of the above equation, resulting in the following 'transfer' function:

$$\frac{Z(\omega)}{X(\omega)} = \frac{1}{1 + G |N(\omega)|^2} \quad (4)$$

where $Z(\omega)$, $X(\omega)$, and $N(\omega)$ are the Fourier transforms of $z(t)$, $x(t)$, and $n(t)$, respectively. If the jammer spectrum is sharply peaked at certain frequencies, the processor serves to suppress these components by an inverse filtering operation.

Space Integrating (SI) Optical Processor

A. Description

An acoustooptic convolver can be implemented using the SI architecture shown in Fig.3. The specific set up shown in Fig.3 is a coherent version of a geometry which can just as easily be implemented using incoherent techniques. The light diffracted from AOD1 is imaged onto AOD2, resulting in two counter-propagating signals at the exit plane of AOD2. These field amplitudes are then spatially integrated by lens L3, and the integrated signal is detected by the photodetector to yield the photocurrent, with a component proportional to:

$$I_1(t) = (2/T) \int_{-T/4}^{T/4} i_1(t+\tau) i_2(t-\tau) d\tau, \quad T=W/v, \quad \tau=x/v, \quad (5)$$

where v is the acoustic velocity in the AODs, and W is the aperture of the AOD. A change of the integration variable leads to a more familiar form given by

$$I_1(t) = (2/T) \int_{t-T/2}^{t+T/2} i_1(\tau) i_2(2t-\tau) d\tau, \quad (6)$$

which is recognized to be a finite window convolution, time-compressed by a factor of two. The complication of this time compressed output becomes an advantage when considering the implementation of the correlator.

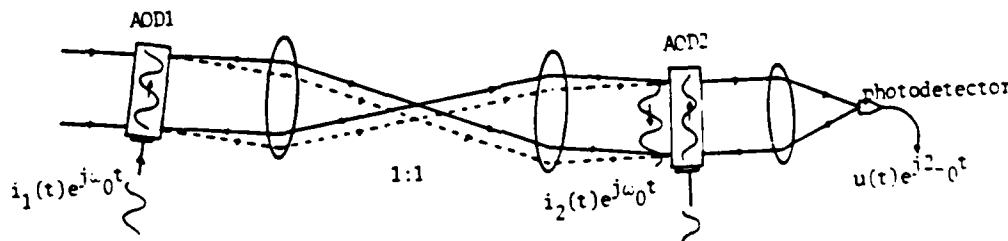


Fig.3 Space Integrating Convolver

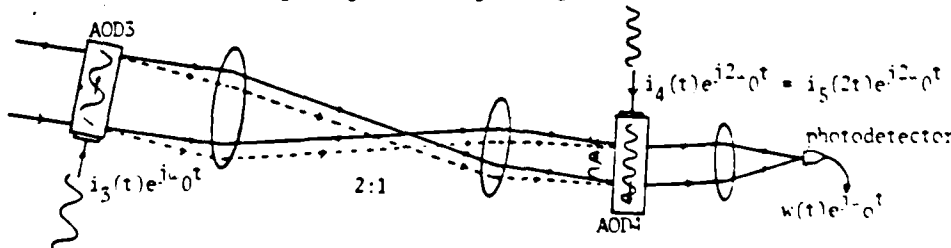


Fig.4 Space Integrating Correlator

Correlation can be performed by the set up shown in Fig.3 if one of the input signals is time-reversed. Since this is incompatible with real-time operation, a different geometry, one that is compatible with the time-compressed output of the convolver, is examined. Correlation requires that the input signals slide past each other without coordinate inversion. This is achieved by the implementation shown in Fig.4 in which the diffracted light from the first AOD is imaged with a 2:1 demagnification onto the second AOD. At the exit plane of the second AOD, the two acoustic signals are co-propagating with different velocities because of the demagnification. These signals are then spatially

integrated onto a single detector. If the input signal to the second AOD is time compressed by a factor of two, i.e., $i_4(t) = i_3(2t)$, then the correlator output is

$$I_2(t) = (2/T) \int_{-T/4}^{T/4} i_3^*(t+2\tau) i_4(t+\tau) d\tau = (2/T) \int_{t-T/2}^{t+T/2} i_3^*(\tau) i_3(t+\tau) d\tau. \quad (7)$$

The above is the desired finite window correlation of the two signals, $i_3(t)$ and $i_5(t)$. Since the output of the convolver in Fig.3 is indeed time-compressed by a factor of two, it can be used to drive the second AOD of the correlator, resulting in the cascaded correlation-convolution operation that is required in the passive processor.

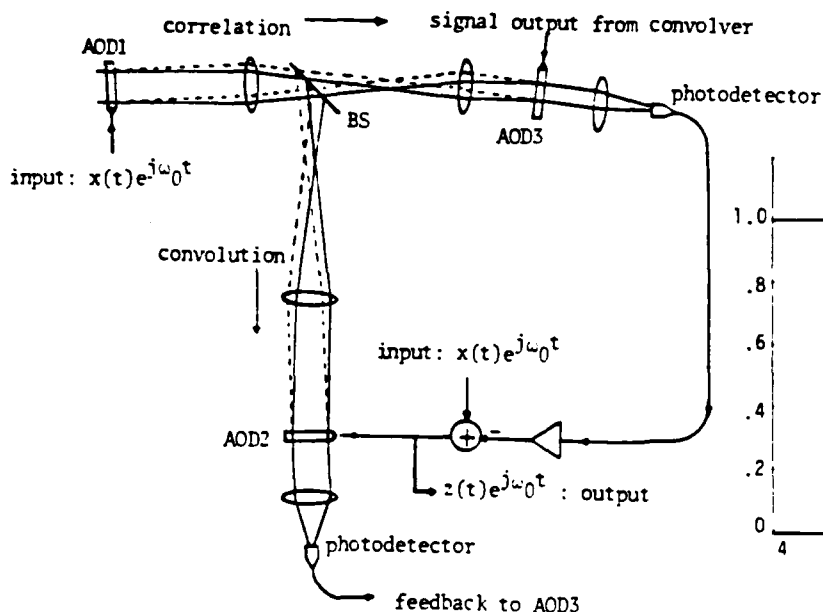


Fig.5 Space Integrating Adaptive Optical Processor

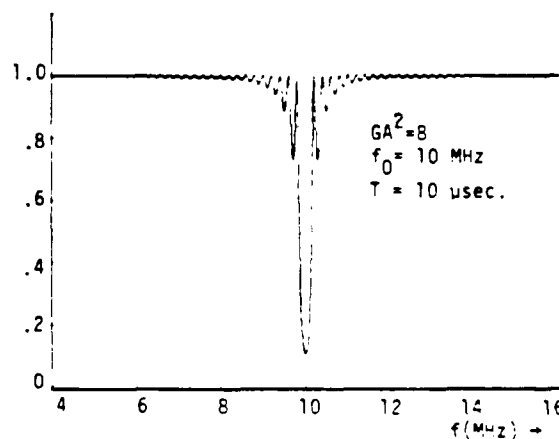


Fig.6 Freq. Response to One Jammer

Shown in Fig.5 is the complete processor. Since the correlator and convolver have one common input, one AOD can be shared by the two blocks, resulting in a system requiring only 3 devices. The input-output equation describing the processor in Fig.5 is given by

$$z(t) = x(t) - (4G/T^2) \iint_{-T/4}^{T/4} x^*(t+2v) x(t+\tau+v) z(t-\tau+v) d\tau dv, \quad (8)$$

where $x(t)$ is the input to the system and $z(t)$ the output. For $x(t) = s(t) + n(t)$, where $n(t)$ is the narrowband jammer and $s(t)$ the broadband signal, the above equation can be Fourier transformed to yield

$$\frac{Z(\omega)}{X(\omega)} = \frac{1}{1 + (G/\pi^2) \int_{-\infty}^{\infty} |N(a)|^2 \text{sinc}^2[(\omega-a)T/4\pi] da} \quad (9)$$

The above 'transfer' function has the form of the inverse filter described earlier, except that the noise spectrum is smoothed by a sinc function due to the finite time windows of the AODs.

B. Performance

The optical system described thus far was implemented for preliminary study. Jammer suppression was successfully observed; however, the maximum, stable null depth obtained was limited to 7dB. It is important to theoretically characterize the performance limits of the processor in terms of system parameters such as the AOD aperture size and detector noise before we attempt to systematically improve the performance level. Three important

issues are discussed: 1) the resolution with which the processor discriminates against jammers, 2) null depth limitations due to detector noise, and 3) transients in the system response due to the delays inherent in the system.

The resolution of the system is characterized entirely by the size of the AOD apertures for they determine the length of the signal sample at any one time. When the input consists of a single sinusoid at frequency ω_0 and amplitude A , the exact form for the Fourier transform of the output can be determined from Eq.8 to be

$$Z(\omega) = \frac{X(\omega)}{1 + GA^2 \text{sinc}^2[(\omega - \omega_0)T/4\pi]} \quad (10)$$

where $Z(\omega)$ and $X(\omega)$ are the Fourier transforms of the output and input signals, respectively. A plot of $Z(\omega)$ for $GA^2 = 8$, $T = 10 \mu\text{secs.}$, and $\omega_0/2\pi = 10 \text{ MHz}$ is shown in Fig. 6. The sidelobe structure is entirely due to the finite apertures of the AODs. The sidelobes can be reduced at the cost of widening the null at the jammer frequency by introducing apodization functions through the appropriate illumination of the AODs. The issue of apodization will be examined more closely later in the discussion of transients and stability.

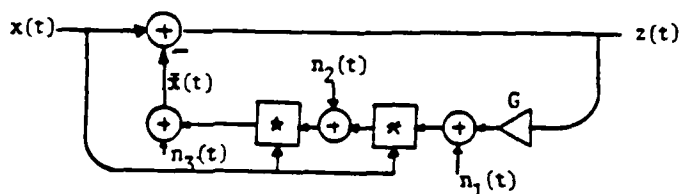


Fig. 7 System Noise Model

While the finite apertures determine the resolution of the processor, the maximum nulling depth is governed by other mechanisms. In order to better understand this limitation, the deterioration in the cancellation of one jammer, $n(t) = A \exp(j\omega_0 t)$, due to the system noise contributed by the detectors and amplifiers is now studied. The system model depicting the noise sources is shown in Fig. 7, where $n_1(t)$ is the noise from the feedback amplifier while $n_2(t)$ and $n_3(t)$ correspond to those arising from the photodetectors in the convolver and correlator, respectively. The noise processes are assumed to be independent, stationary, complex gaussian processes, each with variance σ^2 . Since the input is a single frequency jammer, the convolver and correlator behave as narrow-bandpass filters centered at that frequency. If the space bandwidth products of the convolver and correlator are larger than the feedback gain, the contribution of the noise sources $n_1(t), n_2(t)$, filtered by the convolver and correlator is negligible compared to $n_3(t)$ which is added directly onto the feedback signal $\tilde{x}(t)$. The portions of noise which are fed back are suppressed by the correlator and convolver and can be neglected to first order. Thus, only the noise due to $n_3(t)$ appears directly at the output. We expect that the performance limit set by the system noise would not be reached unless the jammer amplitude is suppressed to σ , the noise level. In our experiments, however, the jammer in the output signal was clearly stronger than the system noise, showing that the noise performance limit was not reached.

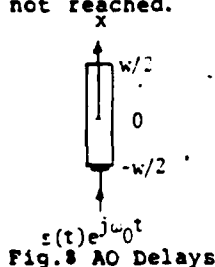


Fig. 8 AO Delays

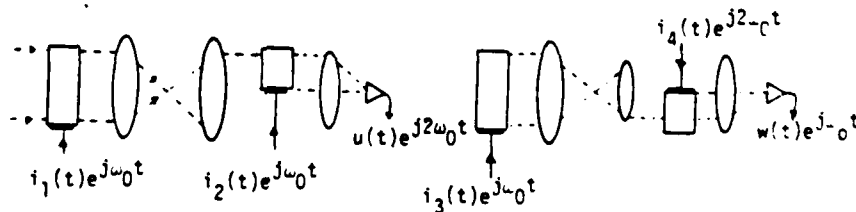


Fig. 9 Delay Modified Convolver and Correlator

A more severe limitation can arise from the time delays that exist in the AODs. Shown in Fig. 8 is an AOD with the spatial coordinate, x , centered on the AOD. If an electrical signal $s(t)$ is used to drive the AOD, then the amplitude of the acoustic wave in the AOD is proportional to $s(t - x/v - T/2)$, where T is the total aperture time. This effectively introduces time delays to appear in the feedback path which can cause instabilities in the processor. To minimize the delays, the AODs in the convolver and correlator can be shifted in space as described in Fig. 9 to result in the following convolver and correlator input-output equations, including the delays:

$$\text{convolver: } u(t) = (4/T) \int_0^{T/4} i_1(t - T/2 + \tau) i_2(t - \tau) d\tau \quad (11)$$

$$\text{correlator: } w(t) = (4/T) \int_{-T/4}^0 i_3^*(t-T/2+2\beta) i_4(t+\beta) d\beta.$$

The resulting modified system is described by the following equation:

$$z(t) = x(t) - 16(G/T^2) \int_{-T/4}^0 \int_0^{T/4} x^*(t-T/2+2\beta) x(t-T/2+\tau+\beta) z(t-\tau+\beta) d\tau d\beta. \quad (12)$$

For a single jammer input, $x(t) = A \exp(j\omega_0 t)$, the above can be transformed to give the following, exact frequency domain result:

$$Z(\omega) = \frac{X(\omega_0)}{1 + GA^2 \text{sinc}^2[(\omega - \omega_0)T/8\pi] \exp[-j(\omega - \omega_0)T/4]} \quad (13)$$

The phase factor present in the denominator is due to the delays. The consequence of this phase factor is that the denominator is no longer positive for arbitrarily large values of G . In fact, the denominator vanishes for $(\omega - \omega_0)T/4 = \pi$ with $G = G_c = \pi^2/4A^2$, indicating instability for gain values larger than G_c which corresponds to an amplitude suppression of only 10dB. This value of maximum suppression is verified by computer simulations of the processor response to a step input. Fig.10 shows the envelope of the response of the processor to a unit step jammer input, for $G=1$, and $G=2.5$, which is near the critical value, G_c . The initial constant portion of the response is due to the delays in the convolver and correlator. Indeed, the second plot shows an oscillatory behavior, indicating marginal stability.

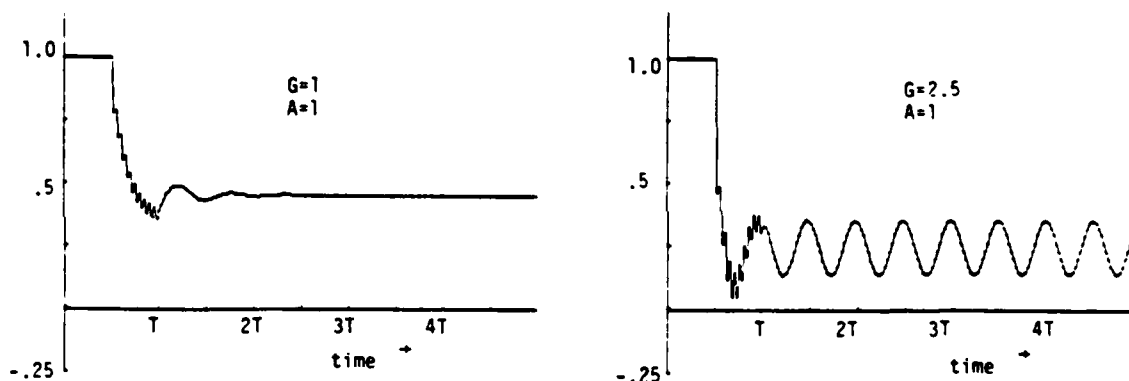


Fig.10 Step Response

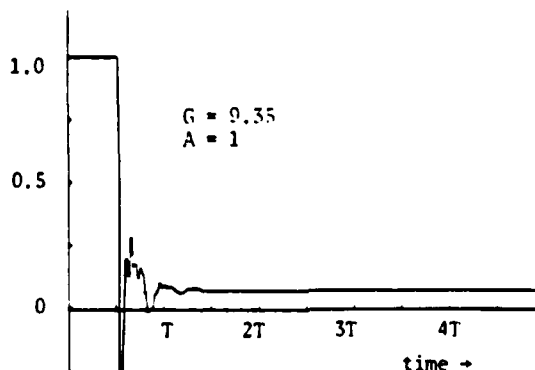


Fig.11 Step Response with Exponential Apodization

The low critical gain stems from the rather large sidelobes of the sinc^2 function coupled with the phase factor due to the delays. At the cost of loss in resolution, the sidelobes can be reduced by apodizing the AOD windows, resulting in systems which can tolerate higher feedback gains while remaining stable. As a simple example, the effect of an exponential tapering function was evaluated by computer simulations. Shown in Fig.11 is the step response obtained with exponential apodization. Convergence is still rapid, and

the suppression is improved to 20dB. Other tapering functions may yield stable nulling performance to match that predicted by the system noise considerations as discussed earlier but at the cost of loss in resolution.

Space-Time Integrating Optical Processor

In the SI system, the desired correlation function is formed by a SI correlator and read out with a SI convolver. Alternative approaches where integration in both time and space are used have been described by Rhodes⁵ and Penn⁶. In such systems, the correlation integral is computed by a time-integrating spatial light modulator (SLM) and read out with a space-integrating convolver. We now consider an implementation using the photorefractive crystal BSO as the time-integrating SLM. The advantages of this implementation over the previous ones^{5,6} which used the Hughes liquid crystal light valve and a phosphor screen, respectively as the time-integrating SLMs are: 1) the system response speed is easily controllable since the response time of the photorefractive effect depends on the exposure level. 2) a coherent implementation is possible. 3) higher resolution is exhibited by the photorefractive devices. 4) since the crystal does not respond to DC or bias levels, the implementation will not suffer from bias build-up problems common among time-integrating systems. 5) the photorefractive devices are simpler since they require only the crystal and a voltage source to supply the external field which is required for some crystals.

The photorefractive effect can be understood in the following way. Two writing beams of light which are spatially modulated with information intersect in the crystal to form an intensity grating pattern. Because of the spatially varying intensity pattern, the charge carriers in the crystal redistribute themselves into a space-charge grating. The electric field associated with this charge grating accordingly modulates the refractive index of the crystal through the electrooptic effect, allowing the read-out of the written information by a third beam. A property characteristic of this effect is that the response time is inversely proportional to the average writing beam intensity impinging on the crystal. This property translates to the controllability of the system response time in the present application, an added flexibility. In our experiments, the time constant was on the order of a few seconds because of the relatively low writing intensities used.

Shown in Fig.12 is a time integrating correlator that exploits the time response characteristics just described. AOD1 and AOD2 serve to spatially modulate the writing Argon laser beams with two counter-propagating signals $s_1(t-x/v)$ and $s_2(t+x/v)$; an imaging lens pair is used to image the AODs onto the BSO crystal. The index grating formed within the crystal is modulated by the correlation of the two signals. The diffracted read-out wave contains only the modulation of the index grating and is thus bias-free. It can be shown that the amplitude of the diffracted light is:

$$E_{\text{diff}}(x) = \int_0^t e^{-(t_1-t)/\tau} s_1(t_1-2x/v-T/2) s_2^*(t_1+2x/v-T/2) dt_1, \quad (14)$$

$$= \int_{t-2x/v-T/2-\tau}^{t-2x/v-T/2} s_1(t_2) s_2^*(t_2+4x/v) dt_2,$$

τ = time constant of BSO.

which is the cross-correlation function of $s_1(t)$ and $s_2(t)$ with the spatial coordinate, x , being the shift variable. This correlation can be detected by imaging the crystal onto a CCD detector array for serial read-out.

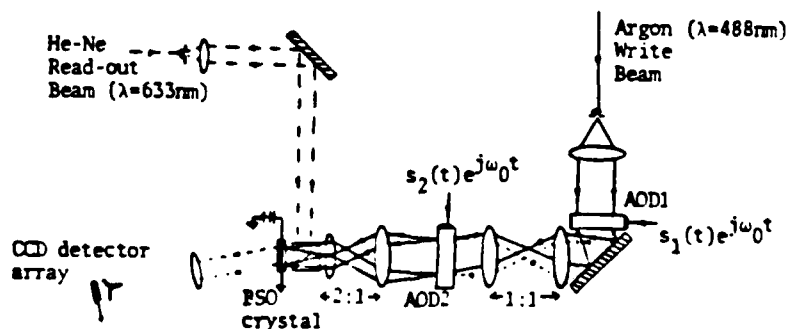


Fig.12 BSO Correlator

The incorporation of the correlator just described into the adaptive system of Fig.2 is rather simple and requires the addition of only 1 AOD with a 4:1 telescope to compensate for the 2:1 demagnification used in writing the correlation onto the BSO. The system is shown schematically in Fig.13. The output and input ($z(t)$ and $x(t)$, respectively) constitute the inputs to the correlator. AOD3 is also driven by the input signal and diffracts a portion of the He-Ne read beam to pass through the BSO unmodulated. The DC from AOD3, however, is Bragg matched to the index grating in the crystal and reads out the correlation function. If all of the AODs are driven at the same frequency, then the beam diffracted by the BSO crystal and that diffracted by AOD3 are collinear and interfere temporally at ω_0 , the Doppler frequency introduced by AOD3. The two diffracted waves are Fourier transformed by lens L_5 , and the resulting intensity is detected by the photodetector. The output current is the convolution of the input, $x(t)$, and the correlation formed inside the crystal, and it becomes the feedback signal, $\bar{x}(t)$, which is then subtracted from the input to produce the output signal. The resulting input-output eqn. is

$$z(t) = x(t) - G \int_{-T/8}^{T/8} \int_{t-\tau}^t x^*(t' - 2a - T/2) x(t - 4a - T/2) z(t' + 2a - T/2) dt' da, \quad (15)$$

which is similar to the SI system except for the presence of both space and time integrals.

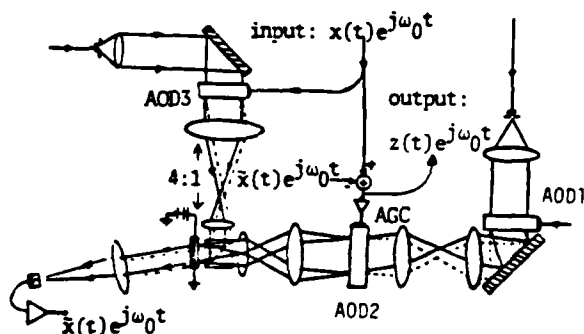


Fig.13 Space-Time Integrating Processor

and $z(t)$ begins to decrease, the AGC tracks the drop in amplitude to keep the input of AOD2 at a fixed value. The system just described has been set up in the laboratory for preliminary study. Shown in Fig.14 is a result showing the spectrum of the input signal, a jammer at 70 MHz, and that of the processor output. The observed suppression was about 15 dB. Better suppression can be expected as better AGC amplifiers are used.

One underlying assumption behind Eq.14 and hence 15 also, is that the crystal is thin enough to put the read out process into the Raman-Nath regime. This is not the case, however, and the coupling between the writing beams must be taken into account. Coupled wave theory predicts that the phase of the index grating will be a function of the ratio of the two write beam intensities; this was found to be the case experimentally. This introduces a complication in the present system because, as soon as the processor begins to suppress an incoming jammer, the output, $z(t)$, begins to decrease in amplitude, causing the ratio of the intensities of the two write beams which emerge from AOD1 and AOD2 to diminish. This translates to a phase change in the feedback signal, possibly causing instability to appear. This can be remedied, however, by placing an AGC (automatic gain control) amplifier at the input to AOD2. As adaptation proceeds

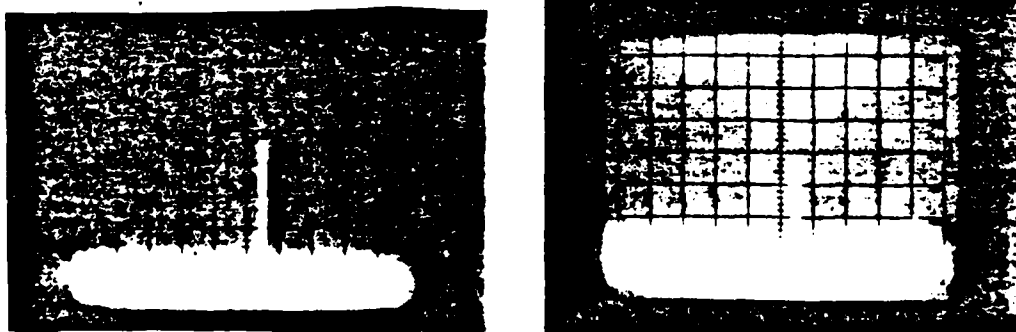


Fig.14 Suppression of Jammer (70 MHz) with BSO Processor
a) input spectrum b) output spectrum

Conclusion

Two optical implementations of adaptive signal processors were described. The space-integrating processor was shown to exhibit very fast response times but also stability problems which could be corrected to a large extent by apodization techniques. The use of such a processor would be appropriate in situations where the jammer noise is varying rapidly in time, such as in the case of blinking jammers or frequency hopped jammers. The space-time integrating system was implemented with a photorefractive BSO SLM to form the necessary correlation function in conjunction with a SI convolver for adaptive filtration. Although considerably slower than its SI counterpart, it has the flexibility of a controllable response time. Future research in photorefractive materials may yield faster crystals.

Acknowledgement

This research is supported by the Air Force Office of Scientific Research and in part by the Rome Air Development Center.

References

1. W. Davenport and W. Root, Introduction to Theory of Random Signals and Noise (McGraw-Hill, New York, 1967).
2. D. Psaltis and J. Hong, 'Adaptive Acoustooptic Filter,' Appl. Opt. 23, 3475 (1984).
3. R. Arens, 'Complex Processes for Envelopes of Normal Noise,' IRE Trans. on Info. Th., IT-3, Sept. 1957.
4. N.V. Kukhtarev, et al, 'Holographic Storage in Electrooptic Crystals, Parts I and II,' Ferroelectrics 22, pgs. 949- , (1979).
5. J. Rhodes, 'Adaptive Filter with a Time-Domain Implementation Using Correlation Cancellation Loops,' Applied Optics 22, 282 (1983).
6. W.A. Penn, et al, 'Acousto-optic Adaptive Processing,' Final Draft, RADC-TR-83-156, Phase Report, Dec. 1983.
7. D. Psaltis, J. Yu, and J. Hong, 'Bias-free Time Integrating Correlator Using a Photorefractive Crystal,' submitted to Applied Optics.
8. A.V. Vanderlugt, 'Adaptive Optical Processor,' Appl. Opt. 21, 4005, (1982).

Measurement of the temporal coherence properties of pulsed single-mode laser diodes

Michael Haney and Demetri Psaltis

The coherence of a single-mode laser diode, under high-speed pulsed modulation, is limited by instabilities in the lasing wavelength arising from transient phenomena in the junction region of the laser. This paper reports the results of an experiment to characterize the effects of these modal instabilities on the temporal coherence of pulsed laser diodes. The primary intent of the experiment was measurement of the cumulative effect of the modal instabilities on the fringe visibility in interferometric time integrating optical processors. A conclusion of this study is that commercially available laser diodes can be used as pulsed light sources in interferometric applications in which the pulse width of the laser is long compared to its characteristic coherence time constant and short compared to its characteristic thermal time constant. Furthermore, the inter-pulse modal instability can be minimized by prudent choice of operating conditions.

I. Introduction

The advancement of the state of the art of commercially available laser diodes in recent years has led to their wide use in optical communication and information processing systems.¹⁻³ Laser diodes are particularly attractive for applications in which volume and power consumption are constrained due to their small size and high efficiency. In addition laser diodes can be modulated directly with bandwidths of up several gigahertz. When laser diodes are pulsed or modulated the spectral purity of the radiation is reduced from that achievable under cw operation, and the temporal coherence decreases accordingly. This paper reports the results of a study to characterize the phenomena which degrade the coherence of single-mode laser diodes under pulsed operation.

A primary factor affecting the coherence of pulsed laser diodes is the rise in temperature of the junction area during the pulse due to ohmic heating.⁴ This effect causes a change in the lasing wavelength and, therefore, degrades the temporal coherence. Analytical models for this behavior, based on the thermodynamic properties of the laser diode structure,^{5,6} are in good agreement with measured results.^{7,8} Other phenomena which affect the temporal coherence of pulsed laser

diodes include the coherence time constant of the laser and mode hopping. In this paper these effects are characterized individually for commercially available single-mode laser diodes, and the overall coherence function resulting from the combination of the effects is modeled and measured.

The coherence time constant of the laser diode is defined as the time after the onset of laser oscillation in which the laser changes from its initial multimode operation to a single dominant lasing mode. The coherence time constant sets a lower bound on the pulse width, since it must be a negligible portion of the pulse width to obtain coherence. After single-mode operation has been established, the stability of the laser mode is affected by junction heating, which leads to a continuous change in the lasing wavelength throughout the duration of the pulse and intrapulse mode hopping which results in an abrupt change in wavelength. When considering time integrating interferometric systems, where many pulses may be integrated, the pulse-to-pulse or interpulse mode stability must also be determined. If the laser does not consistently select the same mode in which to oscillate for every pulse, the coherence will be degraded. Finally, it should be mentioned that the coherence properties of the laser diode are influenced by the operating conditions, and, therefore, measurements were made over a wide range of operating conditions.

Most of the measurements were made on the Hitachi HLP-1600 laser. Measurements were also made on the Mitsubishi ML-3001 and RCA C86030E lasers for comparison. Unless otherwise stated, the data presented are for the Hitachi device. The experimental setup with which the measurements were made is described in the next section. The experimental results are presented as well as the trade-offs that exist between performance and operating conditions. Finally, the

The authors are with California Institute of Technology, Department of Electrical Engineering, Pasadena, California 91125.

Received 26 December 1984.

0003-6935/85/131926-07\$02.00/0.

© 1985 Optical Society of America.

coherence function for the pulsed laser diode is derived and compared with the experimental results.

II. Experiment

The pulsed laser diode coherence measurement setup is shown schematically in Fig. 1. The setup is a Michelson interferometer consisting of mirrors *M1* and *M2* and the beam splitter *B1*. Mirror *M1* is mounted on a micrometer stage to allow variations in optical path difference (OPD) from 0 to 5 cm. Mirror *M2* is mounted on a piezoelectric translator to allow convenient monitoring of the fringe depth at the output of the interferometer. The laser drive electronics consist of a current pulser with variable rise time and dc biasing capability.

The light at the output of the interferometer is detected by a high-speed avalanche photodiode and independently by an integrating linear photodiode array. Beam splitter *B2* permits simultaneous data collection with the two types of detector. The avalanche photodiode that was used is capable of measuring rise times of 1 nsec. The photodiode array temporally integrates the interferometrically detected light of many sequential pulses to characterize the pulse-to-pulse coherence of the laser. When data are being taken with the photodiode array the output beams of the interferometer are slightly misaligned along the dimension of the photodiode array by tilting mirror *M1*. This produces a linear fringe pattern on the array, and the fringe visibility of the interference pattern that is observed on the array is a measure of the coherence over many pulses.

A. Coherence Time Constant

The coherence time constant is estimated by measuring the time from the beginning of detectable light intensity to the time of maximum fringe visibility. This is accomplished by modulating the OPD with mirror *M2* by $\pm\lambda/2$ at ~ 1 kHz with the piezoelectric translator. When many individual pulses are simultaneously displayed on the screen of the oscilloscope, the persistence of the screen's phosphor causes the trace to smear due to the varying relative phase between the interfering beams. This smearing occurs only when the laser is oscillating in a single longitudinal mode. During the start-up period of the pulse, the light is incoherent, and thus the trace of many pulses appears as a single line. The coherence time constant is approximately the difference between the time light is first detected as a single trace and the time of maximum smearing due to interference.

Figure 2 shows the interferometric activity at the leading edge of the laser pulse. The first 8 nsec of a 50-nsec pulse with 15-mW peak power are shown. The OPD for the data of Fig. 2 is 1 cm, and the time scale of the displayed data is 1 nsec/div. The envelope of the trace in Fig. 2 is a measure of the modulation depth of the interference, and thus it provides an estimate of the coherence time. The coherence time constant in Fig. 2 is ~ 1.5 nsec. Prebiasing the laser at a dc level just below the threshold current of the laser reduces the coherence time constant slightly but not sufficiently to

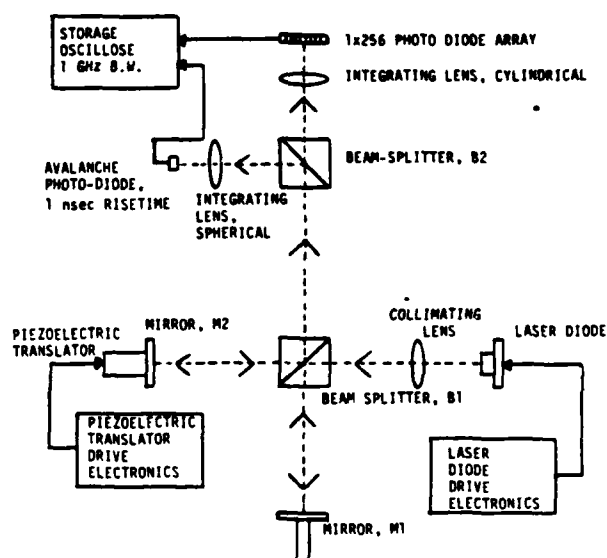


Fig. 1. Laser diode temporal coherence measurement setup.



Fig. 2. Coherence time measurement, 1 nsec/div.

allow accurate measurements of the effect at the short (subnanosecond) time scale involved.

The coherence time constants of the Mitsubishi and RCA lasers were also measured. These lasers have coherence time constants comparable with those of the Hitachi laser diode. At large peak power (>35 mW), however, the coherence time constant for the RCA laser diode became dramatically longer. In this case, the shape of the interferometrically detected light pulse indicated that two or more competing modes continued to oscillate simultaneously for periods of up to 50 nsec before one became dominant.

In conclusion, it was found that the coherence time constant of the single-mode lasers we tested was between 1 and 2 nsec. Therefore, if the laser pulse is much longer than 2 nsec, the coherence time constant will not significantly affect the coherence properties of the laser. For example, if 50-nsec pulses are used, the laser will be oscillating in a single mode 96–98% of the time.

B. Intrapulse Stability

The intrapulse stability of the lasing mode is measured by studying the trace of the output of the ava-

temperature change on the phase difference measured with the interferometer:

$$d(\Delta\beta)/dT \approx -2\pi d\lambda'/\lambda^2. \quad (5)$$

The change in junction temperature during the beginning of laser oscillation has been modeled as follows:

$$\Delta T(t) = T_0[1 - \exp(-t/\tau)], \quad (6)$$

where T_0 is the asymptotic value of the junction temperature change, and τ is the thermal time constant which is determined by the thermal resistance and capacitance of the junction. By integrating Eq. (6) and substituting in Eq. (7) the following expression for the phase difference as a function of time is obtained:

$$\Delta\beta(t) = 2\pi dT_0\lambda'[1 - \exp(-t/\tau)]/\lambda^2. \quad (7)$$

The change in junction temperature is plotted in Fig. 4 as a function of time from the data in Fig. 3. The phase of the amplitude modulation is used as a measure of the change in junction temperature with time. The data clearly show that the temperature rise is exponential in nature and is a close match to the model of Eq. (7) with time constants that are equal to several hundred nanoseconds.

The rates of wavelength shift of the RCA and Hitachi lasers were compared at an output power of 20 mW for 100-nsec pulse widths. It was found that the wavelength of the RCA laser shifted at approximately one half of the rate of the Hitachi laser. This is attributed to the larger optical cavity of the RCA device which leads to a longer thermal time constant.

Mode hopping is the second type of intrapulse mode instability that was observed. This phenomenon manifests itself as a discontinuity in the intensity measured at the output of the interferometer. Figure 5 illustrates the effect of an intrapulse mode hop in a 20-mW pulse. The OPD, pulse amplitude, and pulse width were chosen sufficiently large that the effect of the junction heating as well as the intrapulse mode hopping were observed. The pulse width in Fig. 5 is 800 nsec, and the OPD is 4 cm. At this particular OPD the sudden change in mode during the pulse manifests itself as a step change in the phase of the intensity modulation of $\sim 180^\circ$ indicating that the OPD is close to a value which is an integral number of $\lambda^2/2(\delta\lambda)$, where $\delta\lambda$ is the intermode spacing of the laser.

In general, the probability of occurrence of an intrapulse mode hop increases at higher pulse amplitude and longer pulse widths. For the Hitachi laser diode it was found that for a pulse width of 50 nsec, the frequency of intrapulse mode hops is $<1/40$ for a peak power level of 12 mW or less. The RCA laser at 12 mW was found to have about twice the rate of mode hopping of the Hitachi laser.

The rates of mode hopping given above are averaged. The actual mode hopping rates varied significantly from measurement to measurement. Intrapulse mode hopping occurs because the positions of the modes shift in relation to the gain curve of the laser as the temperature of the junction region rises during the pulse. If

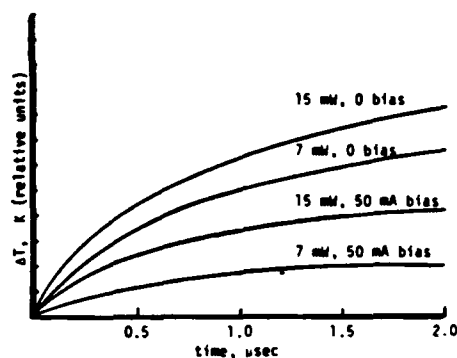


Fig. 4. Rise in junction temperature for different operating conditions.

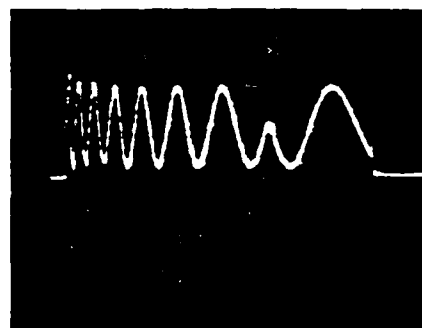


Fig. 5. Intrapulse mode hop, peak power = 20 mW, bias = 0.100 nsec/div.

the shift is large enough, an adjacent mode will move closer to the center of the gain curve than the current lasing mode and will take over. Slight changes in the ambient temperature of the laser's heat sink will also manifest themselves as shifts in the mode position relative to the laser's gain curve thereby changing the mode hopping rate and/or the position within the pulse at which the mode hop occurs.

C. Interpulse Stability

Interpulse instabilities occur when over a long sequence of pulses the laser does not always choose the same mode in which to oscillate. This effect was also characterized in the experimental setup of Fig. 1. The pulse repetition frequency was set at 20 kHz, the pulse width at 50 nsec, and the peak power at 10 mW. The OPD for these data was 1 mm. Under these operating conditions intrapulse mode hopping is negligible. Furthermore, the change in the phase difference between the interfering beams due to junction heating is small enough so that the resulting output can be considered a sampled version of the cw case. The OPD was modulated by $\pm\lambda/2$ at a frequency of 1 kHz. Modulating the OPD in this manner reveals the fringe depth of the interfering beams. An interpulse mode hop will manifest itself as a discontinuity in the sampled waveform. The output of the avalanche photodiode was monitored on the oscilloscope, and an oscilloscope trace of this measurement is shown in Fig. 6. Figure 6 shows

that one mode is dominant, but occasionally an adjacent mode is selected. Different modes will, in general, have different relative phases between the output beams of the interferometer. This causes the detected intensities for the two modes to be different, as shown in Fig. 6. It is clear that for time integrating interferometric systems, in which the light from many laser pulses is detected interferometrically and summed, interpulse mode hopping could severely degrade the fringe visibility and thus must be minimized.

As with the intrapulse mode hopping, the rate of interpulse mode hops was found to be determined by the electrical operating conditions of the laser. In general, the laser was more stable (a single mode was more likely to dominate) at lower peak power levels. This is probably due to the more gradual buildup of laser oscillation at the lower drive current levels resulting in more consistent selection of the dominant mode. It was also found that the rate of mode hopping could be significantly reduced by altering the shape of the current drive pulse to have a slower rise time. Measurements of the interpulse mode hopping rates were made for different drive pulse shapes and peak power levels. For these measurements the slope of the leading edge of the drive pulses was varied while maintaining the peak level pulse width at ~ 50 nsec. Figure 7 shows oscilloscope traces showing the shapes of typical drive pulses that were compared. For the Hitachi laser it was found that for 10-nsec rise times, the interpulse mode stability was $>95\%$ for peak power levels below 6 mW. However, for drive pulses with rise times of 80 nsec, such as in Fig. 7(b), the laser diode exhibited $>95\%$ stability for peak power levels up to 12 mW. For rise times >80 nsec no measurable improvement in the interpulse mode stability was observed. Furthermore, it was noted that applying a dc bias current to the laser at a level below the threshold current did not significantly affect interpulse stability.

D. Coherence Function Measurement

The four types of instability phenomenon that were discussed in the previous sections combine to degrade the fringe visibility in an interferometric system. The relative importance of these instabilities to coherence degradation depends on the operating conditions of the laser diode and the OPD of the optical system. The effect of the coherence time constant is minimal if it is short when compared to the pulse width. For the measurements reported here this condition was generally met, and, therefore, the coherence time constant was a minor factor. Furthermore, we found that intrapulse and interpulse mode hopping are minimized by proper choice of operating conditions.

For example, with the Hitachi laser at a peak power of 10 mW, the combined intrapulse and interpulse mode hopping occurs at a rate of less than 1 mode hop in twenty pulses for a current drive pulse whose shape is similar to that of Fig. 7(b). The most significant instability is the wavelength shift due to the heating of the junction area during the pulse. To maintain a high degree of coherence for a particular OPD, the pulse

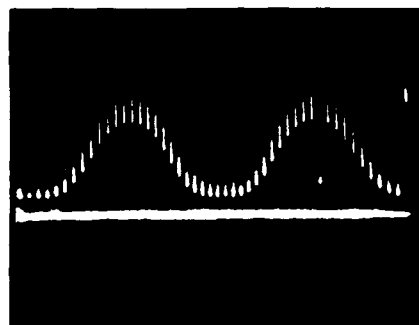


Fig. 6. Interpulse mode hopping.

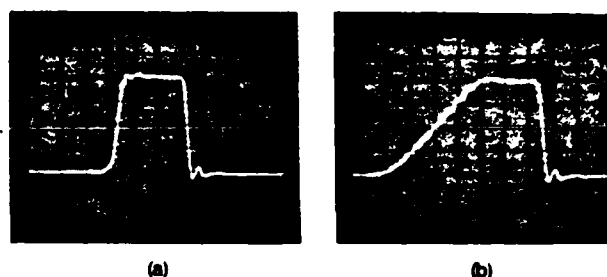


Fig. 7. Laser drive pulses: (a) 10-nsec rise time; (b) 80-nsec rise time.



Fig. 8. Fringe pattern generated by 10,000 interferometrically detected LD pulses.

width must be such that the intensity measured at the output of the interferometer will fluctuate only a fraction of one of the cycles shown in the examples of Fig. 3.

Evaluation of the combined effects of the individual instabilities on the visibility of the interference pattern that results from the integration of many sequential pulses was accomplished by studying the output of the photodiode array. Figure 8 is an oscilloscope trace of the array output. The trace represents the time integrated interferometric response of $\sim 10,000$ individual laser pulses over a period of 50 msec. The data shown were taken at approximately zero OPD and thus represent the largest attainable fringe depth. In the example of Fig. 8 the pulse width is 50 nsec. The actual fringe depth achieved is not 100% because of several imperfections of the system. The photodiode array has dark noise and signal spreading which will lead to bias

contributions. Furthermore, the very short multimode portion at the beginning of each pulse adds incoherently to result in a contribution to the bias.

A measurement of the fringe visibility, normalized to the maximum fringe visibility obtained at OPD = 0, as in Fig. 8, provides a direct measurement of the coherence function under different operating conditions. The measured fringe visibility as a function of OPD for different pulse widths is plotted in Fig. 9. In these data the effects of mode hopping have been minimized, leaving junction heating as the primary factor that is responsible for reduced fringe visibility.

III. Coherence Function Analysis

In this section we derive the coherence function for pulsed laser diodes based on the model of the thermodynamic properties of the laser junction given in Eq. (6). The coherence function analysis is then extended to time integrating systems by including the effect of interpulse mode hopping.

The coherence time constant of a single-mode laser diode has been treated theoretically in the literature.¹¹ It was predicted that, at moderate to high peak power levels, one longitudinal mode dominates very rapidly (~1 nsec) after the onset of laser oscillation. This is in good agreement with our measurements. The coherence time phenomenon made a negligible contribution to the coherence degradation when the pulse width is very long compared to the coherence time constant. We assume this to be the case and neglect this effect in our analysis. We also ignore intrapulse mode hopping because it was also negligible under the operating conditions used. Therefore, the dominant phenomena degrading the coherence of the laser are the intrapulse modal instability due to junction heating and interpulse mode hopping.

For a single pulse, the coherence function of the single-mode laser diode is assumed to be determined solely by the junction heating as modeled by Eq. (6). This effect is deterministic, and, therefore, in this case the coherence function does not involve ensemble averages. The coherence function for multiple pulses, on the other hand, with interpulse mode hopping, is stochastic in nature. We will first derive an expression for the effects of junction heating and then add the effects of random interpulse mode hopping.

The electric field at the output of the interferometer is given by

$$E = \exp[j\omega(t)t] + \exp[j\omega(t - d/c)(t - d/c)], \quad (8)$$

where d is the OPD of the interferometer and $\omega(t)$ is the laser's oscillating frequency which changes with time due to the junction heating effect. From (7) the wavelength as a function of time is given by

$$\lambda(t) = \lambda_i + T_0\lambda'[1 - \exp(-t/\tau)], \quad (9)$$

where λ_i is the lasing wavelength at the start of the pulse. Since the change in wavelength due to the junction heating is small compared with λ_i , we have

$$\omega(t) = 2\pi c/\lambda(t) \approx (2\pi c/\lambda_i)[1 - \lambda'T_0[1 - \exp(-t/\tau)]/\lambda_i] \quad (10)$$

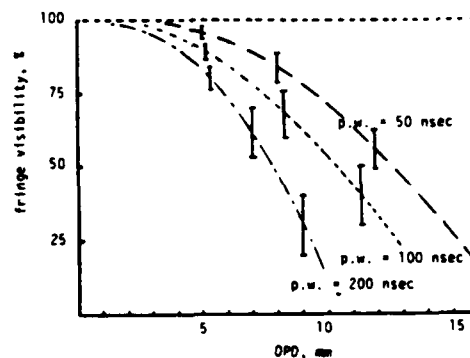


Fig. 9. Measurements of the modulus of the coherence function for 10,000 time integrated pulses.

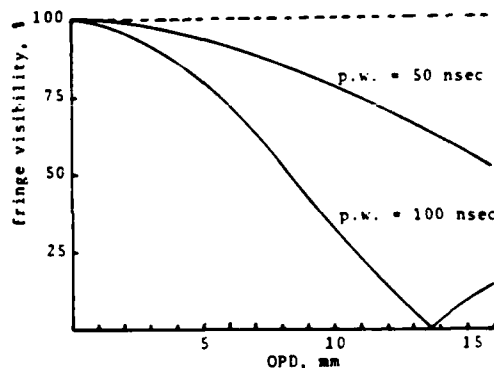


Fig. 10. Theoretical modulus of the coherence function.

We now write an expression for the normalized coherence function as a function of the OPD:

$$G(d) \approx \langle \exp[j\omega(t)t] \exp[-j\omega(t)(t - d/c)] \rangle, \quad (11)$$

where $\langle \rangle$ indicates time average over the pulse duration P . In (11) $\omega(t - d/c)$ is approximated by $\omega(t)$ because $-d/c$ is negligible with respect to τ . We have

$$G(d) = 1/P \int_0^P \exp[j\omega(t)d/c] dt. \quad (12)$$

Substituting in Eq. (10) and using the assumption that $P \ll \tau$, so that $\exp(-t/\tau) \approx 1 - t/\tau$ on the interval P , the integration in Eq. (12) can be performed in closed form:

$$G(d) = \text{sinc}(dc_1P) \exp[j2\pi(d/\lambda_i - dc_1P/2)], \quad (13)$$

where $c_1 = \lambda'T_0/\lambda_i^2\tau$ and $\text{sinc}(x) = \sin(\pi x)/\pi x$.

The fringe visibility V is defined as the magnitude of the coherence function. Hence, for pulse widths much less than the thermal time constant, we have

$$V = |\text{sinc}(dc_1P)|. \quad (14)$$

Equation (14) is plotted in Fig. 10 for pulse widths of 50 and 100 nsec. A thermal time constant of 350 nsec is assumed. The measured fringe visibility (Fig. 9) is in good agreement with Fig. 10. Note that the effect of a large pulse amplitude on the coherence function would

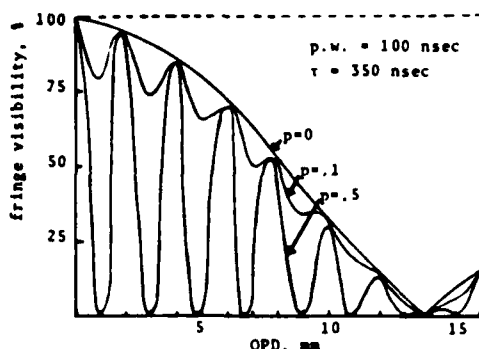


Fig. 11. Calculated effect of interpulse mode hopping on the modulus of the coherence function.

be to increase T_0 and hence c_1 in Eq. (14), thus leading to lower coherence.

We now analyze the fringe visibility for the case of two competing adjacent longitudinal modes. The two modes have frequencies $\omega_1(t)$ and $\omega_2(t) = \omega_1(t) + \delta\omega$, where $\delta\omega$ is the Fabry-Perot mode spacing of the laser diode. The time dependence is assumed to be due solely to the junction heating effect. We assign p and $1 - p$ to be the probabilities of occurrence of the two modes. For any one pulse, only one of the two modes exists, and thus the coherence function for multiple time integrated pulses is

$$G_2(d) = E\{(\exp[j\omega(t)t] \exp[-j\omega(t)(t - d/c)])\}. \quad (15)$$

Using Eq. (13) we find

$$G_2(d) = \exp[j(\omega_{1,i}d/c - \pi dc_1P)] \text{sinc}(dc_1P) \times [p + (1 - p) \exp[j(\delta\omega)d/c]], \quad (16)$$

where we have ignored the insignificant differences in the phase term and in the argument of the sinc function for the two modes. $\omega_{1,i}$ and $\omega_{2,i} = \omega_{1,i} + \delta\omega$ are the initial lasing frequencies for two modes. Hence the fringe visibility is given by

$$V_2 = |\text{sinc}(dc_1P)| [2p^2 - 2p + 1 + 2p(1 - p) \cos[(\delta\omega)d/c]]^{1/2}. \quad (17)$$

This function is plotted in Fig. 11 for $p = 0$, $p = 0.1$, and $p = 0.5$ for a pulse width of 100 nsec. Figure 11 depicts the degradation in coherence due to the interpulse mode hopping instability.

References

1. D. Botez and G. Herskowitz, "Components for Optical Communications Systems: A Review," *Proc. IEEE* 68, No. 6 (June 1980).
2. J. G. Duthie, J. Upatnieks, C. R. Christensen, and R. D. McKenzie, Jr., "Real Time Optical Correlation with Solid State Sources," *Proc. Soc. Photo-Opt. Instrum. Eng.* 231, 281 (1980).
3. D. Psaltis, "Optical Image Correlation Using Acoustooptic and Charge-Coupled Devices," *Appl. Opt.*, 21, 491, (1982).
4. Y. A. Bykovskii *et al.*, "Method for Investigating Thermal Condition in and Spectral Characteristics of a Semiconductor Laser by Means of a Fabry-Perot Resonator," *Sov. Phys. Semicond.*, 5, 435 (1971).
5. J. Butler, "The Effect of Junction Heating on Laser Linearity and Harmonic Distortion," in *Topics in Applied Physics*, Vol. 34 (Springer-Verlag, Berlin, 1980), Chap. 8.
6. Y. A. Bykovskii *et al.*, "Coherence of the Radiation of a Pulsed Single Mode Injection Semiconductor Laser," *Sov. Phys. Dokl.*, 17, 359 (1972).
7. P. Melman and W. J. Calsen, "Interferometric Measurement of Thermal Coherence and Time-Varying Longitudinal-Mode Wavelengths in GaAs Diode Lasers," in *Technical Digest, Conference on Lasers and Electro-optics* (Optical Society of America, Washington D. C., 1981), paper WS4.
8. M. Haney and D. Psaltis, "Coherence Properties of Pulsed Laser Diodes," *Proc. Soc. Photo-Opt. Instrum. Eng.* 422, 197 (1983).
9. A. Yariv, *Quantum Electronics* (Wiley, New York, 1975).
10. *Hitachi Laser Diode Application Manual*.
11. K. Y. Lau, C. Harder, and A. Yariv, "Longitudinal Mode Spectrum of Semiconductor Lasers Under High-speed Modulation" *IEEE J. Quantum Electron.* QE-20, No. 1, 71 (Jan. 1984).

The research reported in this paper is supported by the Air Force Office of Scientific Research, General Dynamics, and the Army Research Office.

ACOUSTO-ELECTRO-OPTIC LIGHT MODULATION

Demetri Psaltis, Hyuk Lee and Gabriel Sirat

California Institute of Technology
Pasadena, CA 91125Abstract

The acousto-electro-optic (AEO) interaction that is presented in this paper is based on the combined Acousto-optic (AO) and Electro-optic (EO) effects. The effect is analyzed and demonstrated experimentally. The application of AEO interaction to light modulation and deflection are discussed.

I. Introduction

The most widely used methods for wideband light modulation are either the acousto-optic⁽¹⁾ or the electro-optic⁽²⁾ effects. Each type of modulator suffers from its own distinct limitations. For instance in an AO modulator there is a tradeoff between bandwidth and diffraction efficiency. A small transducer provides a large bandwidth but at the expense of reduced diffraction efficiency and vice-versa. An EO modulator in an anisotropic crystal has low angular aperture, requires high voltage and it is configured with polarizers and a phase retarder requiring tedious alignment. When a modulator is implemented using the combined A-O and E-O effects a new flexibility is introduced⁽³⁾ with which we can overcome some of the limitations of the two individual modulators.

In this paper the A-E-O interaction is introduced (Section 2) and experimentally confirmed (Section 3). An assessment of the properties of an AEO modulator is presented in Section 4 and in Section 5 we discuss the use of the A-E-O effect as a light deflector.

II. Acousto-electro-optic interaction

The basic configuration of an AEO modulator is shown in Fig. 1. The device is operated in a normal configuration of an anisotropic acousto-optic modulator. An acoustic wave is launched in the crystal by the piezoelectric transducer and the device is illuminated at the Bragg angle. In addition, electrodes are deposited on the crystal, so that a high voltage can be applied and thus an electric field introduced in the crystal, at the same time.

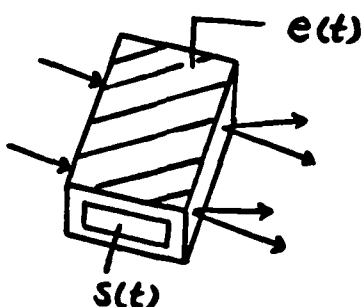


Figure 1. Basic configuration of the AEO device.

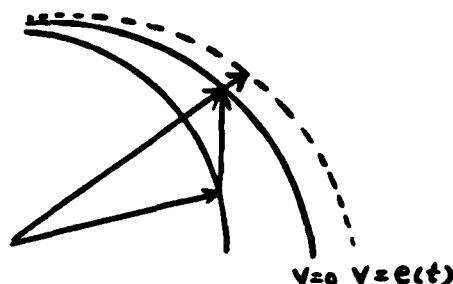


Figure 2. Phase matching diagram of the AEO interaction

The acousto-optic interaction is analyzed by coupled modes analysis⁽⁴⁾ which can be schematically represented by a phase matching diagram (fig. 2). The solid curves in the figure represent the wave-normal surfaces of the crystal without the external voltage. When an external voltage is applied to the two electrodes of Fig. 1, the shape of the index ellipsoid changes due to the E-O effect and the wave-normal surfaces move accordingly (dotted curves in Fig. 2). For simplicity we depict the case that only the outer curve changes. Assuming that the crystal is aligned for perfect Bragg matching for zero applied voltage, then by applying the voltage a phase mismatch is introduced and the diffracted light intensity varies according to the amplitude of the external voltage.

The intensity of the light diffracted in the first order for phase mismatch equal to ΔK is given by⁽⁴⁾

$$(1) \quad I_d = I_{in} \eta [\text{sinc} (\sqrt{\eta + (\Delta K L / 2)^2}]^2$$

where

I_{in} - incident light intensity

η - diffraction efficiency of the acousto-optic interaction

L - interaction length

If we define the voltage V^{AEO} as a measure of the electrically induced phase mismatch in the AEO effect, satisfy the following

$$(2) \quad \frac{\Delta K L}{2} = \frac{\pi V}{V^{AEO}},$$

then V^{AEO} is given by:

$$(3) \quad V^{AEO} = \frac{\lambda}{n^3 \Delta n} \frac{h}{L},$$

where λ -wavelength, n -refractive index, h -height of the crystal Δn the relevant combination of electro-optic coefficients and V is the applied voltage. V^{AEO} is approximately equal to both the half-wave voltage of the A-E-O modulator and to the half voltage of the same device utilized as an E-O transverse modulator. We can rewrite (1) in terms of V^{AEO} as follows:

$$(4) \quad I_d = I_{in} \eta [\text{sinc} (\sqrt{\eta + (\frac{\pi V}{V^{AEO}})^2}]^2$$

This is the diffracted light intensity formula of the AEO interaction and we can see that this is a direct effect, i.e., the modulation function, in Eq. (4), is not simply the product of the electro-optic and acousto-optic effects.

III. Experiments

An experimental device was constructed to verify the A-E-O interaction. (A photograph of the device is shown in Fig. 3). The device was constructed from a Y-cut LiNbO₃ crystal with dimensions 40x7x12 (x-y-z) mm. Two electrodes were evaporated on the (0,1,0) surfaces and connected to a high voltage power supply. An acoustic transducer, mounted on the (1,0,0) surface, launches a transverse, y-polarization, acoustic wave in the x-direction. A beam from a He-Ne laser propagates near the z-axis (1.8° from z-axis). The intensity of the diffracted light (normalized to ηI_{in}) is plotted in Fig. 4 as function of the voltage (normalized to V^{AEO}/π). The theoretical (solid) curve is given by Eq. (1) and (4).

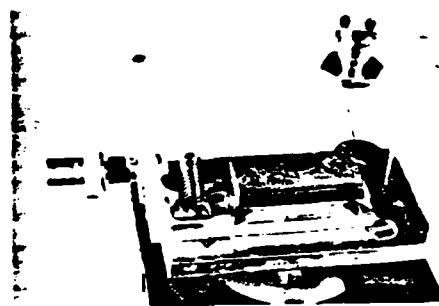


Figure 3. Photograph of the experimental device.

In this experiment V^{AEO} was 6 kV and η , the diffraction efficiency was 2%.

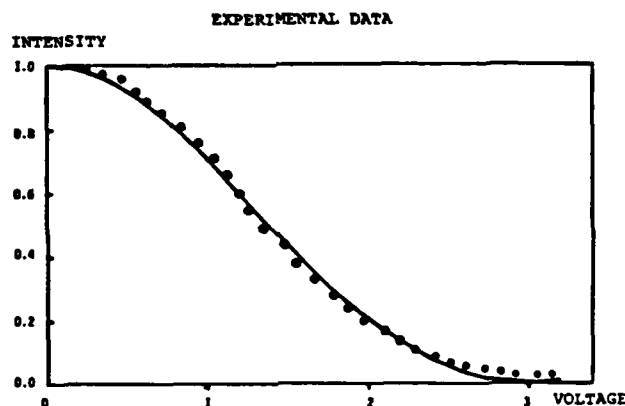


Figure 4. Experimental results of an AEO modulator.

—: Theoretical Curve $INTENSITY = [SINC(VOLTAGE)]^2$
 O : Experimental Data, $VOLTAGE = \pi V / V^{AEO}$
 $V^{AEO} = 6 \text{ kV}$

The experimental results are in excellent agreement with the theoretical curve.

IV. A-E-O modulator

The simplest device that can be implemented with the AEO interaction is a light modulator, in the arrangement of the experiment discussed in the previous section. Specifically if the acoustic signal consists of a single fixed frequency (the center frequency of the device) with constant power, then the diffracted light intensity can be modulated by varying the voltage applied across the electrodes of the device, in accordance to the curve in Fig. 4. Interestingly, the same device can be used as either an acousto-optic modulator by temporally modulating the acoustic power or alternatively as an electro-optic modulator by varying the applied voltage and observing the transmitted light through crossed polarizers. In this section we compare the AEO modulator to the properties of the two individual modulators. The bandwidth of the AEO modulator is limited primarily by the electrical driving circuit with which the voltage is applied across the electrodes.

The piezoelectric transducer does not affect, in any way the bandwidth and therefore it can be designed to maximize the diffraction efficiency by increasing the transducer (and thus the interaction) length.

In addition the acoustic transit time limitation that is associated with the AO modulator is not encountered here and thus an extended optical beam can be modulated. Therefore the temporal characteristics of the AEO modulator are similar to the characteristics of the EO modulator which are generally superior to those of the AO modulator. The spatial characteristics of the AEO modulator, on the other hand, are the same as the AO modulator. The modulated light is angularly separated from the unmodulated light and thus the need for cross polarizers is eliminated. The AEO modulator works with unpolarized light as well (with a loss in diffraction efficiency by a factor of 2) and in general the alignment of the AEO (and AO) modulator is less tedious. For instance, the numerical aperture of the AEO device we discussed in the previous section is 1.1° an EO device using the same crystal has only $.6^\circ$ numerical aperture. The interaction geometry of the AEO modulator can be optimized to yield several degrees of numerical aperture.⁽⁵⁾ This is the most important advantage of the AEO modulator over the EO device.

The modulation curve of the AEO modulator (diffracted intensity vs. applied voltage) is given by Eqs. (1) and (4) and is plotted in Fig. 4. This modulation curve is different from the EO or the AO modulators, but the linear dynamic range that can be obtained from the modulator is approximately the same. For instance we have compared the linearity of the AEO and EO modulators by calculating the modulation depth obtained in the two cases when the third harmonic of modulation is set to 1% of the fundamental in both cases. The modulation efficiency in this case is 44% for the AEO modulator and 48% for the EO modulator.

The voltage level that is required in the AEO device is similar to that needed in a transverse EO modulator. The required voltage can be reduced by increasing the interaction length, reducing the height of the device and using materials with higher electro-optic constants. The geometrical improvements can be most conveniently accomplished through the use of surface acoustic waves and optical waveguides. This is a possibility we are currently exploring.

V. Acousto-electro-optic deflector

Another possible use of the AEO interaction will be to build an AEO phase mismatch compensated deflector.

The number of resolvable spots in an AO deflector is given by (1)

$$N = \frac{\Delta\theta_T}{\Delta\theta_0} ; \quad \Delta\theta_0 = \frac{\lambda}{nD} ,$$

where $\Delta\theta_T$ is the total deflected angle, λ is the optical wavelength, n is the index of refraction and D is the optical aperture. $\Delta\theta_T$ is given by the bandwidth of the deflector. This bandwidth is given by the transducer bandwidth and the phase mismatch introduced by the deviation of the scanning acoustic wave frequency from the center frequency. If the appropriate voltage is applied to compensate the phase mismatch introduced for each scanning frequency, a larger bandwidth can be obtained and therefore a larger total deflection angle and a larger number of spots. The compensating voltage is approximately linear in the phase mismatch, ΔK and therefore also linear in the change in frequency $f - f_c$.

The device discussed in section 3 was also used to demonstrate the AEO deflector. The results of these experiments are shown in Fig. 5. The normalized intensity of the diffracted light obtained with and without the compensating voltage is plotted as function of the acoustic frequency. The compensating voltage, as function of acoustic frequency is plotted in the same figure. From Fig. 5 we can see that the bandwidth of the AEO deflector is about 2.5 times larger than that of the AO deflector and is limited by the electrical bandwidth of the transducer. As expected the compensating voltage is linear in $f - f_c$.

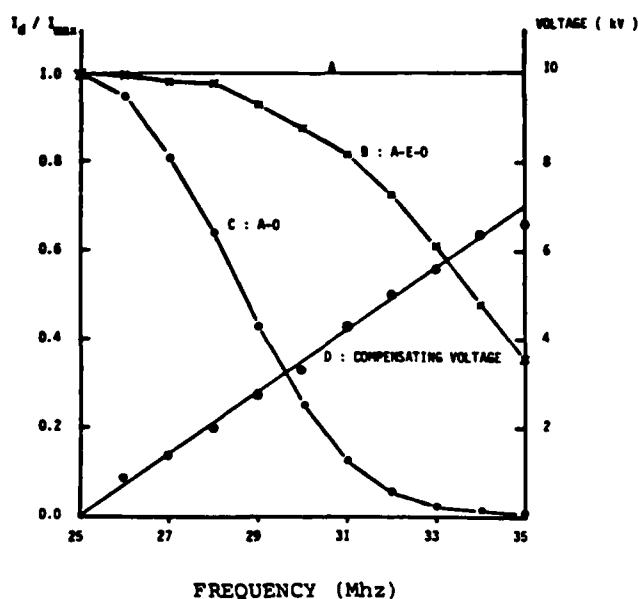


Figure 5. Experimental results of a phase-mismatch compensated AEO deflector

- A. Ideal response of the Compensating Deflector.
- B. Real deflected light intensity of the Compensating Deflector due to the transducer bandwidth.
- C. Deflected light intensity without compensating voltage.
- D. Compensating Voltage.

VI. Conclusion

The most important characteristic of the AEO effect is the increased flexibility in the design that results from the additional modulating signal. This can allow the design to overcome some of the limitations imposed by the individual EO and AO interactions.

The AEO effect can prove to be of particular significance in the area of optical signal processing where the direct interaction among the three input signals to the device (optical, acoustic, electrical) can allow signal processing operation to be efficiently performed.

The principal limitation of the AEO effect is the high voltage requirement in our present device. We are exploring several methods for reducing the required voltage including the use of planar technology (surface acoustic waves and optical waveguides)⁽⁷⁾ and the resonant piezoelectric effect.⁽⁶⁾

Acknowledgement

The research reported in this paper is supported in part by the Air Force Office of Scientific Research.

References

1. I. C. Chang, IEEE Trans. Sonics Ultrason. SU-23,2 (1976).
2. A. Yariv, "Introduction to Optical Electronics," (Holt, Reinhart and Winston, 1976).
3. James J. Wiczer and Henry Merkelo, Appl. Phys. Letters 30, 439 (1977).
4. J. M. Rouvaen, M. G. Ghazaleh, E. Bridoux and R. Torquet, J. Appl. Phys. 50(80) 5472, August 1979.
5. I. C. Chang, Appl. Phys. Letters 25, 370 (1974).
6. G. Sirat, Y. Fisher and N. Ben Yosef, Proceeding of SPIE Vol. 465
7. J. M. Hammer, "Modulation and Switching of Light in Dielectric Waveguides," Integrated Optics, Topics in Applied Physics, Vol. 7, pp. 138-200, Springer-Verlag Berlin Heidelberg, New York, 1979.

Acousto-electro-optic light modulation

Demetri Psaltis, Hyuk Lee, and Gabriel Sirat
California Institute of Technology, Pasadena, California 91125

(Received 24 September 1984; accepted for publication 12 November 1984)

The acousto-electro-optic effect is analyzed and demonstrated experimentally. The application of the acousto-electro-optic effect to light modulation and deflection is discussed.

The most widely used methods for wideband light modulation are either the acousto-optic (AO)¹ or the electro-optic (EO)² effects. Each type of modulator has its own strengths and suffers from its own distinct limitations.³ When we use AO and EO effects simultaneously,⁴ the acousto-electro-optic (AEO) effect, a new flexibility is intro-

duced with which we can overcome some of the limitations of the two individual modulators.

The basic configuration of an AEO modulator is shown in Fig. 1. The device is operated in the normal configuration of an anisotropic AO modulator. A single frequency acoustic wave is launched in the crystal by the piezoelectric trans-

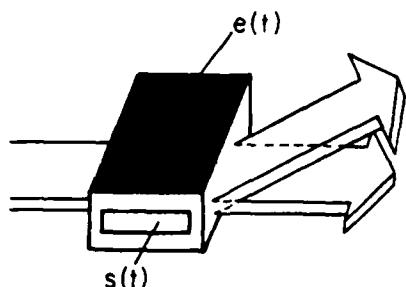


FIG. 1. Basic configuration of the AEO device.

ducer and the device is illuminated at the Bragg angle corresponding to the given frequency. The AO interaction is analyzed by coupled mode analysis.⁵ At the Bragg angle, the diffracted light intensity is maximum. If there is a phase mismatch ΔK , the diffracted light intensity I_d is given by (the approximation $\eta \ll 1$)

$$I_d = \eta I_{in} \{ \text{sinc}[\eta + (\Delta K L / 2)^2]^{1/2} \}^2 \approx \eta I_{in} \{ \text{sinc}(\Delta K L / 2) \}^2, \quad (1)$$

where I_{in} is the incident light intensity, η is the diffraction efficiency of the AO interaction, L is the interaction length, and $\text{sinc}(x) = [\sin(x)]/x$. In the AEO device electrodes are deposited on the crystal, so that an external electrical field can be applied in addition to the acoustic wave. The application of the external field causes a deformation of the index ellipsoid due to the EO effect, and the wave-normal surfaces move accordingly. Consequently, a phase mismatch ΔK is introduced and the diffracted light intensity varies according to the amplitude of the external voltage. The phase mismatch introduced by the anisotropic AEO effect is given by

$$\Delta K \approx \left(\frac{2\pi}{\lambda} \right) \left[\Delta n_2 - \left(\frac{n_1}{n_2} \right) \Delta n_1 \right] \approx \left(\frac{2\pi}{\lambda} \right) \Delta n, \quad (2)$$

where λ is the wavelength of the light in vacuum, n_1 and n_2 are indices of refraction without voltage, Δn_1 and Δn_2 are the changes of n_1 and n_2 due to EO effect, and $\Delta n = \Delta n_2 - \Delta n_1$. The phase mismatch of the isotropic AEO effect is very small compared with that of anisotropic AEO effect because in the isotropic case, $n_1 = n_2$ and $\Delta n_1 = \Delta n_2$, and hence $\Delta K \approx 0$. For the anisotropic case we can write Δn as follows²:

$$\Delta n \approx [(n_1^3 r_1 - n_2^3 r_2) / 2] V / h, \quad (3)$$

where h is the height of the crystal, V is the applied voltage, and r_1 and r_2 are EO coefficients. We define the voltage V^{AEO} as a measure of the electrically induced phase mismatch in the AEO effect, satisfying the following:

$$(\Delta K L / 2) = (\pi V / V^{AEO}). \quad (4)$$

From equations (2)-(4) we have

$$V^{AEO} \approx (\lambda h / L) [(n_1^3 r_1 - n_2^3 r_2) / 2]^{-1}. \quad (5)$$

V^{AEO} is approximately equal to twice the half-wave voltage of the same crystal utilized as a transverse EO modulator. We can rewrite (1) in terms of V^{AEO} as follows:

$$I_d = I_{in} \eta \{ \text{sinc}[\eta + (\pi V / V^{AEO})^2]^{1/2} \}^2. \quad (6)$$

This is the diffracted light intensity formula of the AEO interaction and we can see that this is a direct effect, i.e., the modulation function in Eq. (6) is not simply the product of

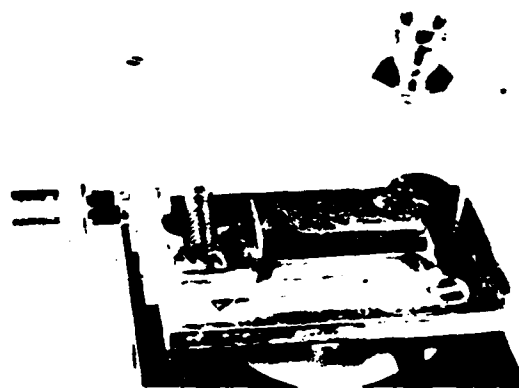


FIG. 2. Photograph of the experimental device. The size of the device was 50 mm \times 20 mm \times 20 mm.

the electro-optic and acousto-optic effects.

An experimental device was constructed to demonstrate the AEO interaction. The device, shown in Fig. 2, was constructed from a Y-cut LiNbO₃ crystal with dimensions 40 \times 7 \times 12 (x - y - z) mm. Two electrodes were evaporated on the (0, 1, 0) surfaces and connected to a high voltage power supply. An acoustic transducer, mounted on the (1, 0, 0) surface, launches a 20-MHz, transverse, y-polarization, acoustic wave in the x direction. The device is illuminated with a collimated beam from a He-Ne laser, polarized in the y direction and propagating at an angle 1.8 (deg) (the Bragg angle) with respect to z axis. The intensity of the diffracted light is plotted in Fig. 3 as function of the voltage (normalized to V^{AEO}/π). The theoretical (solid) curve is given by Eqs. (1) and (6). In this experiment V^{AEO} was 6 kV and η , the diffraction efficiency, was 2%. The experimental results are in excellent agreement with the theoretical curve.

The simplest device that can be implemented with the AEO interaction is a light modulator in the arrangement of the experiment discussed in the above paragraph. Specifically if the acoustic signal consists of a single fixed frequency (the center frequency of the device) with constant power, then the diffracted light intensity can be modulated by varying the applied voltage across the electrodes of the device, in accordance to the curve in Fig. 3. The bandwidth of the AEO modulator is limited primarily by the maximum power supplied by the electrical driving circuit with which the voltage is applied across the electrodes. The voltage level required for an AEO modulator is, within a geometrical factor close

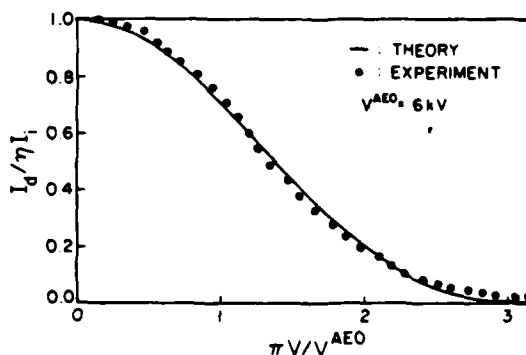


FIG. 3. Experimental results of an AEO modulator.

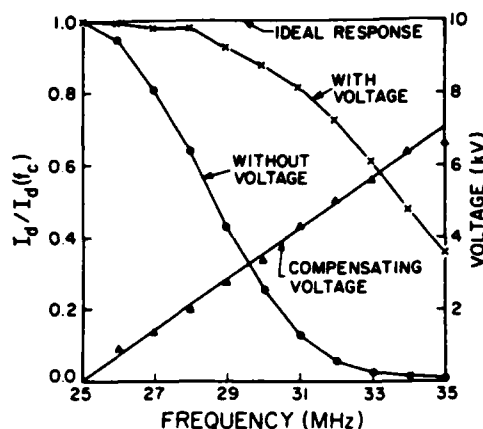


FIG. 4. Experimental results of a phase mismatch compensated AEO deflector.

to unity, twice that needed in a transverse EO modulator of the same geometry. The modulated light is angularly separated from the unmodulated light and the need for cross polarizers is eliminated. The AEO modulator can be operated with unpolarized light with a loss of diffraction efficiency of 2. The alignment of an AEO modulator (like an AO modulator) is almost insensitive to the direction perpendicular to the interaction plane. This is to be compared to the small numerical aperture, in both directions, of an EO modulator, limited by natural birefringence. The piezoelectric transducer does not affect, unlike in the AO case, the bandwidth of the modulator. Therefore, it can be designed to maximize the diffraction efficiency by increasing the transducer width and thus the interaction length. In addition the acoustic transit time limitation that is associated with the AO modulator is not encountered here and thus an extended collimated optical beam can be modulated. The acoustic port can be used to dynamically align the modulator in the plane of the interaction, by changing the center frequency, and to compensate for fluctuations in the optical input intensity by changing the intensity of the acoustic power. We have compared the linearity of the AEO and EO modulators by calculating the modulation depth obtained in the two cases when the third harmonic is set to 1% of the fundamental. The modulation efficiency is 44% in the AEO modulator and 48% in the EO modulator.

AEO interaction can also be used to improve the performance of an AO deflector. The number of resolvable spots in an AO deflector is given by

$$N = \Delta\vartheta_T / \Delta\vartheta_0; \quad \Delta\vartheta_0 = \lambda / nD, \quad (7)$$

where $\Delta\vartheta_T$ is the total deflected angle, λ is the optical wavelength, n is the index of refraction, and D is the optical aperture. $\Delta\vartheta_T$ is limited by the bandwidth of the deflector and determined by the transducer bandwidth and the phase mismatch introduced by the deviation of the scanning acoustic wave frequency from the center frequency. If the appropriate voltage is applied to compensate the phase mismatch introduced for each scanning frequency, a larger total deflection angle and a larger number of spots can be obtained. The phase mismatch Δk is approximately linear in the applied voltage and for small acoustic frequency, which is the case of our experiment, it is also linear in the change of frequency $f - f_c$. Therefore, the compensating voltage is linear in the change of frequency $f - f_c$. The number of spots in an AEO deflector neglecting the influence of the transducer bandwidth is given by

$$N' = N(1 + 2.25V_{\max}/V^{\text{AEO}}), \quad (8)$$

where V_{\max} is the maximum voltage amplitude. The device, shown in Fig. 2, was also used to demonstrate the AEO deflector. The results of this experiment are shown in Fig. 4. The normalized intensity of the diffracted light obtained with and without the compensating voltage is plotted as a function of the acoustic frequency. The compensating voltage, as a function of the acoustic frequency, is plotted in the same figure. From Fig. 4 we can see that the bandwidth of the AEO deflector is about 2.5 larger than that of the AO deflector and was limited by the electrical bandwidth of the transducer. The compensating voltage is approximately linear in $f - f_c$. If the acoustic frequency is sufficiently high, so that the acoustic wavelength is comparable to the optical wavelength, we have to include a quadratic term in the relation between the phase mismatch and the frequency change; therefore, the compensating voltage will also include a quadratic term in this case.

The research reported in this paper was supported by the Air Force Office of Scientific Research.

¹I. C. Chang, IEEE Trans. Sonics Ultrason. SU-23, 2 (1976).

²A. Yariv, *Introduction to Optical Electronics* (Holt, Reinhart and Wiston, NY, 1976).

³D. Psaltis, H. Lee, and G. Sirat, Proceedings of the International Society for Optical Engineering, Los Angeles, 26-27 January 1984, Vol. 465, pp. 171-175.

⁴J. J. Wiczer and H. Merkelo, Appl. Phys. Lett. 30, 439 (1976).

⁵J. M. Rouaven, M. G. Ghazaleh, E. Bridoux, and R. Torquet, J. Appl. Phys. 50, 5472 (1979).

Photorefractive incoherent-to-coherent optical converter

Y. Shi, D. Psaltis, A. Marrakchi, and A. R. Tanguay, Jr.

Y. Shi and D. Psaltis are with California Institute of Technology, Division of Engineering & Applied Science, Pasadena, California 91125; the other authors are with University of Southern California, Departments of Electrical Engineering and Materials Science, Optical Materials & Devices Laboratory, Los Angeles, California 90089. Received 17 September 1983.

Sponsored by William T. Rhodes, Georgia Institute of Technology.

0003-6935/83/233665-03\$01.00/0.

© 1983 Optical Society of America.

Photorefractive materials have been extensively used in recent years as real-time recording media for optical holography.^{1,2} One prospective application of real-time holography is in the area of optical information processing; for example, the correlation between two mutually incoherent images has recently been demonstrated in real time in a four-wave mixing geometry.³ Often, however, the information to be processed exists only in incoherent form. High performance spatial light modulators⁴ are thus necessary in many optical information processing systems to convert incoherent images to coherent replicas for subsequent processing. We report in this Communication the successful demonstration of real-time incoherent-to-coherent image transduction through the use of holographic recording in photorefractive crystals. Several possible configurations and experimental results are presented.

The interference of two coherent beams in the volume of a photorefractive crystal generates nonuniformly distributed free carriers, which are redistributed spatially by diffusion and/or drift in an external applied field. The subsequent trapping of the free carriers in relatively immobile trapping sites results in a stored space-charge field, which in turn modulates the index of refraction through the linear electrooptic effect.⁵ Thus a volume phase hologram is recorded. If the two coherent beams are plane waves, a uniform phase grating is established. An incoherent image focused in the volume of the photorefractive material will spatially modulate the charge distribution stored in the crystal. This spatial modulation can be transferred onto a coherent beam by reconstructing the holographic grating. The spatial modulation of the coherent reconstructed beam will then be a negative replica of the input incoherent image. The holographic

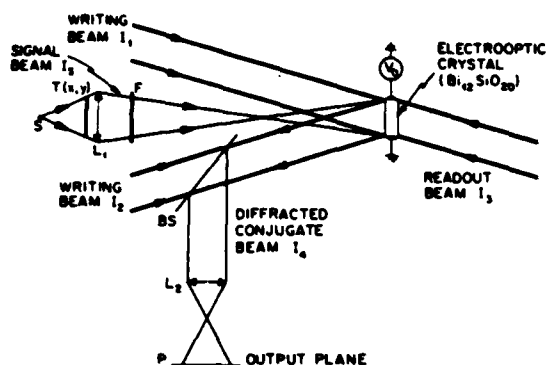
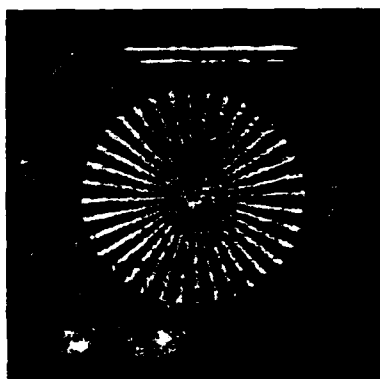


Fig. 1. Experimental setup for incoherent-to-coherent conversion with phase conjugation in four-wave mixing. The writing beams I_1 and I_2 and the reading beam I_3 are generated from an argon laser ($\lambda = 514$ nm). The phase conjugate beam I_4 is diffracted at the same wavelength. The transparency $T(x,y)$ is illuminated with a xenon arc lamp S and imaged on the BSO crystal with the optical system L_1 through a filter $F(\lambda = 545$ nm). BS is a beam splitter, and P is a polarizer placed in the output plane.



(a)



(b)

Fig. 2. Incoherent-to-coherent conversion utilizing phase conjugation in four-wave mixing: (a) spoke target; (b) USAF resolution target. The group 3.6, corresponding to a resolution of 14.3 lp/mm, is well resolved.



Fig. 3. Photorefractive incoherent-to-coherent conversion of a transparency with grey levels.

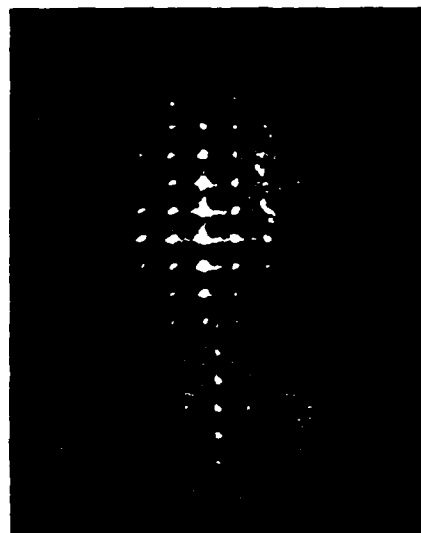


Fig. 4. Fourier transform of a grid pattern formed after a photorefractive incoherent-to-coherent conversion of the grid pattern image.

grating can be recorded before, during, or after the crystal is exposed to the incoherent image. Therefore, a number of operating modes are possible. These include the Grating Erasure Mode (GEM), the Grating Inhibition Mode (GIM), and the Simultaneous Erasure Writing Mode (SEWM), among others.

In the Grating Erasure Mode, a uniform grating is recorded by interfering the two writing beams in the photorefractive crystal. This grating is then selectively erased by incoherent illumination of the crystal with an image-bearing beam. The incoherent image may be incident either on the same face of the crystal as the writing beams or on the opposite face. When the absorption coefficients of the writing and image-bearing beams give rise to significant depth nonuniformity within the crystal, these two cases will have distinct wavelength-matching conditions for response optimization.

In the Grating Inhibition Mode, the crystal is pre-illuminated with the incoherent image-bearing beam prior to grating formation. This serves to selectively decay (enhance) the applied transverse electric field in exposed (unexposed) regions of the crystal. After this pre-exposure, the writing beams are then allowed to interfere within the crystal, causing grating

formation with spatially varying efficiency due to vast differences in the local effective applied field. This technique will also work in the diffusion limit with no external applied field by means of a similar physical mechanism.

In the Simultaneous Erasure Writing Mode, the conventional degenerate four-wave mixing geometry is modified to include simultaneous exposure by an incoherent image-bearing beam, as shown schematically in Fig. 1. Diffraction by a phase grating in the four-wave mixing configuration has been modeled following two different approaches.^{6,7} Common to both analyses, the diffracted intensity is proportional to both the readout intensity and the square of an effective modulation ratio, in the first-order approximation, and assuming no pump depletion. In addition, a uniform beam incident on the photosensitive medium at an arbitrary angle decreases the modulation ratio and hence the overall diffraction efficiency.⁸ In the SEWM configuration, these effects can be combined with the diffraction of a conjugate beam in a four-wave mixing geometry to perform the incoherent-to-coherent image conversion. In particular, this conversion can be regarded as caused by selective spatial modulation of the grating by spatial encoding of the incoherent erasure beam. It should be noted here that a related image encoding process could be implemented in a nonholographic manner by premultiplication of the image with a grating.⁹

In our experiments, we have successfully produced incoherent-to-coherent conversions in all three operating mode configurations as well as in several modifications of the basic arrangements described above. We present here experimental results from our implementations of the Simultaneous Erasure Writing Mode.

The experimental arrangement in one implementation is as shown in Fig. 1. The two plane wave writing beams (labeled I_1 and I_2) are generated from an argon laser ($\lambda = 514$ nm) and interfere inside the photorefractive crystal to create a phase volume hologram. The readout beam I_3 , collinear with I_1 to satisfy the Bragg condition, diffracts the phase conjugate beam I_4 at the same wavelength and with increased diffraction efficiency when a transverse electric field is applied to the electrooptic medium. An incoherently illuminated transparency $T(x,y)$ with intensity $I_s(x,y)$ (either quasi-monochromatic or white light) is imaged in the plane of the crystal. The beam splitter BS separates the diffracted signal from the writing beam; the Polaroid filter P in the output plane eliminates the unwanted scattered light to enhance the signal-to-noise ratio.¹⁰ The photorefractive material utilized was a single crystal of bismuth silicon oxide (BSO), cut to expose polished (110) faces, and of dimensions $7.3 \times 6.9 \times 1.3$ mm³.

A transverse electric field $E_0 = 4$ kV/cm was applied along the $[110]$ axis perpendicular to the polished faces. The carrier frequency of the holographic grating, $f = 300$ lp/mm, was within the optimum range for the drift-aided charge transport process.¹¹ The vertically polarized coherent writing beam and signal intensities were $I_{1,2} = 0.4$ mW/cm² and $I_s = 8$ mW/cm², respectively. Figure 2 shows the converted images obtained from two binary transparencies (a spoke target and a USAF resolution target). The illumination was provided by a xenon arc lamp through a broadband filter centered at $\lambda = 545$ nm (FWHM = 100 nm). An approximate resolution of 15 lp/mm was achieved without optimizing factors such as the optical properties and quality of the crystal, the depth of focus in the bulk of the medium, the carrier frequency of the grating, and the relative intensities and wavelengths of the various beams. This spatial bandwidth is comparable with that obtained with a PROM¹² or a liquid crystal light valve.¹³

The image shown in Fig. 3 demonstrates the capability of

the technique to reproduce many grey levels. To obtain this image, a negative transparency was illuminated with blue light ($\lambda = 488$ nm) derived from an argon laser and was focused on the BSO crystal. The holographic grating was recorded with green light ($\lambda = 514$ nm) and read out with an auxiliary red beam ($\lambda = 6328$ Å).

The 2-D Fourier transform formed by a lens after the incoherent-to-coherent conversion of a grid pattern is shown in Fig. 4. The fundamental spatial frequency of the grid was ~ 1 lp/mm. The existence of several diffracted orders and the well-focused diffraction pattern are positive indications that the device is suitable for coherent optical processing operations.

Although these results are preliminary, they clearly demonstrate the feasibility of real-time incoherent-to-coherent conversion utilizing phase conjugation in photorefractive BSO crystals. This device has potential for incoherent-to-coherent conversion with high resolution, which can be realized by optimizing the optical properties and quality of the crystal, the depth of focus in the bulk of the medium, the carrier frequency of the grating, and the relative intensities and wavelengths of the various beams. In addition, such a device is quite attractive from considerations of low cost, ease of fabrication, and broad availability. With such a device, numerous optical processing functions can be directly implemented that utilize the flexibility afforded by the simultaneous availability of incoherent-to-coherent conversion and volume holographic storage.

The authors would like to thank F. Lum for his technical assistance and Y. Owechko, J. Yu, and E. Paek for helpful discussions. This research was supported in part at USC by RADC under contract F19628-83-C-0031, the Defense Advanced Research Projects Agency, the Joint Services Electronics Program, and the Army Research Office and at Caltech by the Air Force Office of Scientific Research and the Army Research Office.

References

1. P. Gunter, "Holography, coherent light amplification, and optical phase conjugation," *Phys. Rep.* 93,200 (1983).
2. D. M. Pepper, *Opt. Eng.* 21, 156 (1982).
3. J. O. White and A. Yariv, *Appl. Phys. Lett.* 37, 5 (1980).
4. A. R. Tanguay, Jr., in *Proceedings, ARO Workshop on Future Directions for Optical Information Processing, Lubbock, Tex., May, 1980* (1981), pp. 52-77.
5. D. L. Staebler and J. J. Amodei, *J. Appl. Phys.* 43, 1042 (1972).
6. J. Feinberg, D. Heiman, A. R. Tanguay Jr., and R. W. Hellwarth, *J. Appl. Phys.* 51, 1297 (1980).
7. N. V. Kukhtarev, V. B. Markov, S. G. Odulov, M. S. Soskin, and V. L. Vinetskii, *Ferroelectrics* 22, 961 (1979).
8. W. D. Cornish and L. Young, *J. Appl. Phys.* 46, 1252 (1975).
9. R. Grousson and S. Mallick, *Appl. Opt.* 19, 1762 (1980).
10. J. P. Herriau, J. P. Huignard, and P. Aubourg, *Appl. Opt.* 17, 1851 (1978).
11. J. P. Huignard, J. P. Herriau, G. Rivet, and P. Gunter, *Opt. Lett.* 5, 102 (1980).
12. Y. Owechko and A. R. Tanguay, Jr., *Opt. Lett.* 7, 587 (1982).
13. P. Aubourg, J. P. Huignard, M. Hareng, and R. A. Mullen, *Appl. Opt.* 21, 3706 (1982).

Physical characterization of the photorefractive incoherent-to-coherent optical converter

A. Marrekchi

A. R. Tanguay, Jr.

University of Southern California
Departments of Electrical Engineering
and Materials Science
Optical Materials and Devices Laboratory
Los Angeles, California 90089-0483

J. Yu

D. Psaltis

California Institute of Technology
Department of Electrical Engineering
Pasadena, California 91125

Abstract. Volume holographic storage in photorefractive $\text{Bi}_{12}\text{SiO}_{20}$ (BSO) crystals is utilized to perform dynamic incoherent-to-coherent image conversion by means of selective spatial erasure of a uniform grating with white (or quasi-monochromatic) light. Physical characterization of device performance is described, with emphasis on resolution, sensitivity, and temporal response. The photorefractive incoherent-to-coherent optical conversion process is analyzed in terms of a simple model that relates the diffracted intensity to the space-variant effective modulation ratio.

Subject terms: optical information processing components; spatial light modulators; bismuth silicon oxide; volume holographic recording; incoherent-to-coherent conversion; photorefractive devices.

Optical Engineering 24(1), 124-131 (January/February 1985).

CONTENTS

1. Introduction
2. Photorefractive incoherent-to-coherent optical converter
 - 2.1. Principles of operation
 - 2.2. Device resolution considerations
 - 2.3. Theoretical model
 - 2.4. Experimental results
 - 2.5. Response time considerations
3. Conclusion
4. Acknowledgments
5. References

1. INTRODUCTION

Real-time holography in photorefractive materials has been utilized for many applications in the areas of optical processing,¹ nondestructive testing of vibrating structures,² and image propagation through aberrating media.³ In addition, the development of high performance spatial light modulators has received much recent attention, primarily for incoherent-to-coherent conversion functions in optical information processing and computing applications.^{4,5} We have previously reported⁶ the successful application of volume phase holography to the real-time conversion of an incoherent image (quasi-monochromatic or white-light illumination) into a coherent replica. Several possible configurations for the performance of this image transduction in photorefractive materials were discussed therein. Such a conversion process has also been investigated independently.⁷

In this paper, we present a simple model that describes the incoherent-to-coherent conversion process in terms of the space-variant effective modulation ratio. The model is shown to satisfactorily explain the principal features of the device sensitivity behavior. The limiting resolution of the process is discussed, and geometric, systemic, and materials-related constraints are identified. Experimental measurements of the device temporal response characteristics are presented.

Invited Paper IP-107 received July 3, 1984; revised manuscript received Oct. 1, 1984; accepted for publication Oct. 9, 1984; received by Managing Editor Oct. 18, 1984. This paper is a revision of Paper 465-12 which was presented at the SPIE conference on Spatial Light Modulators and Applications, Jan. 26-27, 1984, Los Angeles, CA. The paper presented there appears (unrefereed) in SPIE Proceedings Vol. 465 © 1985 Society of Photo-Optical Instrumentation Engineers.

The high sensitivity of photoconductive and electrooptic crystals such as bismuth silicon oxide ($\text{Bi}_{12}\text{SiO}_{20}$, or BSO) in the visible spectrum has allowed the simultaneous recording and reading of volume holograms to be achieved with time constants amenable to real-time operation. The holographic recording process in photorefractive materials involves photoexcitation, charge transport, and trapping mechanisms.^{8,9} When two coherent beams are allowed to interfere within the volume of such a crystal, free carriers are nonuniformly generated by absorption and are redistributed by diffusion and/or drift under the influence of an externally applied electric field. Subsequent trapping of these charges generates a stored space-charge field, which in turn modulates the refractive index through the linear electrooptic (Pockels) effect and thus records a volume phase hologram. If both coherent writing beams are plane waves, the induced hologram will consist of a uniform grating.

In the photorefractive incoherent-to-coherent optical conversion (PICOC) process, in addition to the uniform holographically induced grating stored in the crystal, an incoherent image is focused in the volume of the photorefractive material and creates a spatial modulation that can be transferred onto a coherent readout beam by reconstructing the holographic grating. The spatial modulation of the coherent reconstructed beam will then be a negative replica of the input incoherent image. The holographic grating can be recorded before, during, or after the crystal is exposed to the incoherent image. Therefore, a number of distinct operating modes are possible, including, among others, the grating erasure mode (GEM), the grating inhibition mode (GIM), and the simultaneous erasure/writing mode (SEWM).

In the *grating erasure mode*, a uniform grating is first recorded by interfering the two coherent writing beams in the photorefractive crystal. This grating is then selectively erased by incoherent illumination of the crystal with an image-bearing beam. The incoherent image may be incident either on the same face of the crystal as the writing beams or on the opposite face. When the absorption coefficients at the writing and image-bearing beam wavelengths give rise to significant depth nonuniformity within the crystal, these two cases will have distinct wavelength-matching conditions for response optimization.

In the *grating inhibition mode*, the crystal is illuminated with the incoherent image-bearing beam prior to grating formation. This

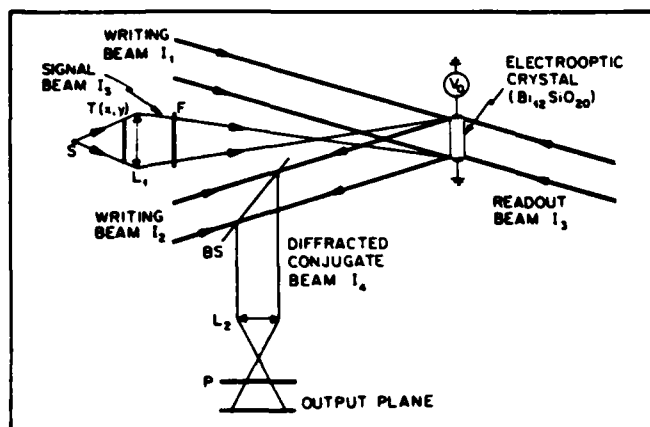


Fig. 1. Experimental configuration for photorefractive incoherent-to-coherent optical conversion (PICOC) in the simultaneous erasure/writing mode. The writing beams I_1 and I_2 and the reading beam I_3 are generated from an argon laser ($\lambda = 514$ nm). The phase-conjugate beam I_4 is diffracted at the same wavelength. The transparency $T(x, y)$ is illuminated with a xenon arc lamp S and imaged onto the $\text{Bi}_{12}\text{SiO}_{20}$ crystal with the optical system L_1 through a filter F ($\lambda = 545$ nm). A coherent replica is formed in the output plane by utilizing beam splitter BS in conjunction with lens L_2 to image the surface of the crystal onto the output plane through polarizer P .

serves to selectively decay (or enhance) the applied transverse electric field in exposed (or unexposed) regions of the crystal. After the incoherent exposure, the coherent writing beams are then allowed to interfere within the crystal, causing grating formation with spatially varying efficiency due to significant differences in the local effective applied field.

In the *simultaneous erasure/writing mode*, the incoherent image modulation, the coherent grating formation process, and the readout function are performed simultaneously. In one possible implementation of this mode, the conventional degenerate four-wave mixing geometry can be modified to include simultaneous exposure by an incoherent image-bearing illumination, as shown schematically in Fig. 1. In this configuration, the coherent replica of the incoherently exposed image appears in the conjugate diffracted beam.

The PICOC process thus can be regarded as being caused by selective spatial modulation of a grating by spatial encoding of an incoherent erasure beam. It should be noted here that a related image encoding process could be implemented in a nonholographic manner by premultiplication of the image with a grating.¹⁰

2. PHOTOREFRACTIVE INCOHERENT-TO-COHERENT OPTICAL CONVERTER

2.1. Principles of operation

The three operating modes described above exhibit significant differences with respect to sensitivity, response time, contrast ratio, and operational complexity. Of these, the simultaneous erasure/writing mode offers the highest degree of operational simplicity due to its inherent nondestructive readout capability and its lack of beam sequencing requirements. In the remainder of this paper, this mode is described in more detail as an example of the essential features of PICOC device operation. Extension of the model to the grating erasure mode and grating inhibition mode can be made by proper incorporation of sequencing time constants. As a specific example of the simultaneous erasure/writing mode, we will discuss the configuration in which the coherent writing beams and the incoherent image-bearing beam illuminate the same face of a BSO photorefractive crystal, with a transverse electric field applied along the $\langle 110 \rangle$ axis, as shown in Fig. 2.

Consider the incident intensity distribution to be composed of two coherent writing beams with intensities I_1 and I_2 and an incoherent signal beam with intensity $I_5(x)$ at the surface of the crystal. The

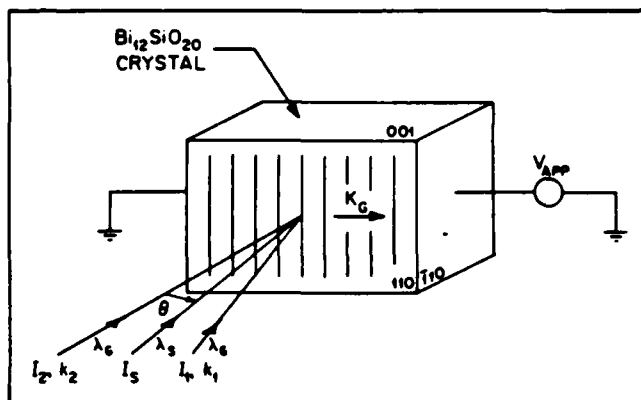


Fig. 2. PICOC transverse electrooptic configuration and recording geometry. The volume holographic grating with wave vector K_G is formed by the coherent writing beams I_1 and I_2 , and the incoherent information is encoded on the beam I_5 .

coherent writing beams will combine to form an interference pattern $I_G(x)$ such that the total incident intensity distribution $[I_T(x)]$ can be written as

$$I_T(x) = I_G(x) + I_5(x) \\ = I_G(1 + m_G \cos K_G x) + I_5(1 + m_S \cos K_S x), \quad (1)$$

where $I_G = I_1 + I_2$ and $m_G = 2(I_1 I_2)^{1/2} / I_G$. K_G and K_S are the wave vectors of the coherent grating (G) and a particular harmonic component of the incoherent image (S), and m_G and m_S are the respective modulation ratios. Due to the nonlinear dependence of the induced space-charge field on the incident intensity, such a distribution will generate a refractive index modulation that can be written in the form

$$\Delta n(x) = C_1 \cos K_G x + C_2 \cos K_S x + C_3 \cos[(K_G - K_S)x] \\ + C_4 \cos[(K_G + K_S)x] + \text{other terms.} \quad (2)$$

Note in particular that $\Delta n(x)$ contains the sum and difference spatial frequencies as well as the fundamental harmonics. The constants C_i ($i = 1$ to 4) represent the amplitudes of each frequency component. Thus, the far-field diffraction pattern shown in Fig. 3 will comprise several beams, each corresponding to a specific term in the refractive index modulation.

To demonstrate incoherent-to-coherent conversion in the four-wave mixing configuration, the experimental arrangement of Fig. 1 was utilized with a 1.3 mm thick (110) cut BSO crystal. A transverse electric field ($E_0 = 4$ kV/cm) was applied along the $\langle 110 \rangle$ axis, and the carrier frequency of the holographic grating ($f = 300$ line pairs/mm) was chosen to lie within the optimum range for drift-aided charge transport in BSO.¹¹

In this implementation, the two plane wave writing beams (labeled I_1 and I_2 in Fig. 1) are generated from an argon laser ($\lambda_G = 514$ nm) and interfere inside the BSO crystal to create a volume phase grating. The readout beam I_3 , collinear with I_1 to satisfy the Bragg condition, is diffracted to form the phase conjugate beam I_4 at the same wavelength. Increased diffraction efficiency results when a transverse electric field is applied to the electrooptic medium. An incoherently illuminated transparency $T(x, y)$, with intensity $I_5(x, y)$ (either quasi-monochromatic or white light), is imaged in the plane of the crystal. The beam splitter BS separates the diffracted signal from the writing beam, and the polarizing filter P in the output plane eliminates the unwanted scattered light to enhance the signal-to-noise ratio.¹² The vertically polarized coherent writing

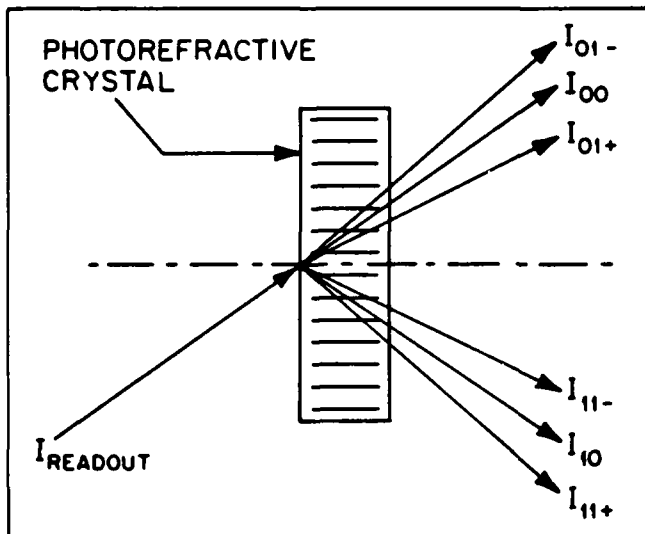


Fig. 3. Far-field diffraction pattern resulting from the nonlinear interaction of two gratings in a photorefractive crystal.

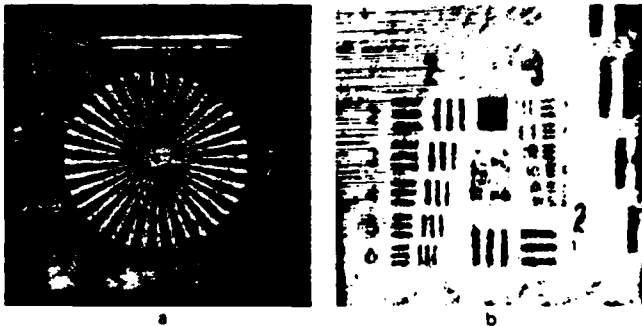


Fig. 4. Photorefractive incoherent-to-coherent optical conversion of two binary transparencies: (a) spoke target and (b) U.S. Air Force resolution target.



Fig. 5. Photorefractive incoherent-to-coherent optical conversion of a gray-level transparency.

beam and signal intensities were $I_{1,2} = 0.4 \text{ mW/cm}^2$ and $I_5 = 8 \text{ mW/cm}^2$, respectively. The incoherent illumination was provided by a xenon arc lamp through a broadband filter centered at $\lambda_5 = 545 \text{ nm}$ (FWHM = 100 nm).

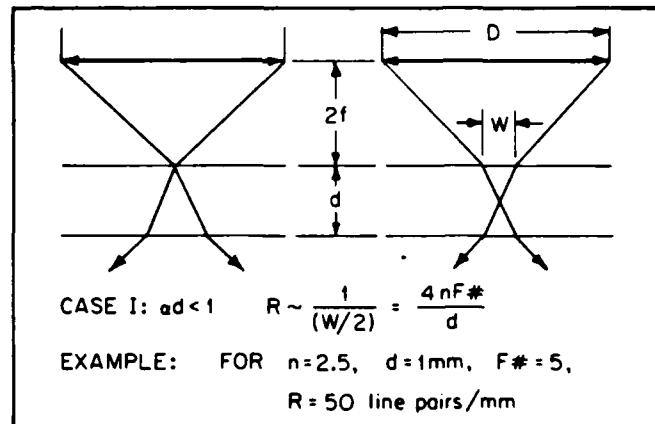


Fig. 6. Geometric constraints imposed on PICOC resolution by the finite f-number of the input optics for the case $ad < 1$.

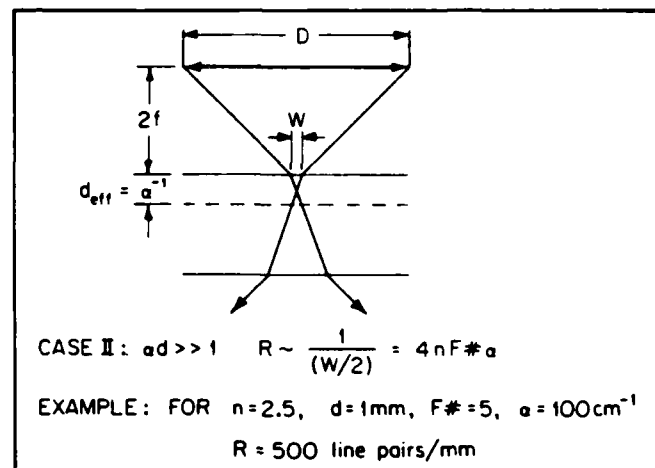


Fig. 7. Geometric constraints imposed on PICOC resolution by the finite f-number of the input optics for the case $ad \gg 1$.

The converted images obtained from two binary transparencies (a spoke target and a U.S. Air Force resolution target) and from a black-and-white slide with gray levels are shown in Figs. 4 and 5, respectively. The original transparency and its converted image have reversed contrast. An approximate resolution of 15 line pairs/mm (as derived from the resolution target image) was achieved without optimizing factors such as the optical properties and quality of the crystal, the depth of focus in the bulk of the medium, the carrier frequency of the grating, the relative intensities and wavelengths of the various beams, or the Bragg diffraction condition (discussed below). This spatial bandwidth is comparable to that obtained with a Pockels Readout Optical Modulator¹³ or a liquid crystal light valve.¹⁴

2.2. Device resolution considerations

A number of distinct factors influence the ultimate resolution achievable with the PICOC spatial light modulator. These factors can be classified as geometric, systemic, or materials-related in nature and are described in detail below.

Geometric resolution limitations derive principally from the incorporation of an incoherent imaging system in the four-wave mixing geometry and from the finite crystal thickness required to create a volume holographic grating. These effects are illustrated in Figs. 6 and 7. Figure 6 describes the case for which $\alpha_0 d \ll 1$ at the grating recording wavelength, such that the induced holographic grating has essentially uniform amplitude throughout the volume of

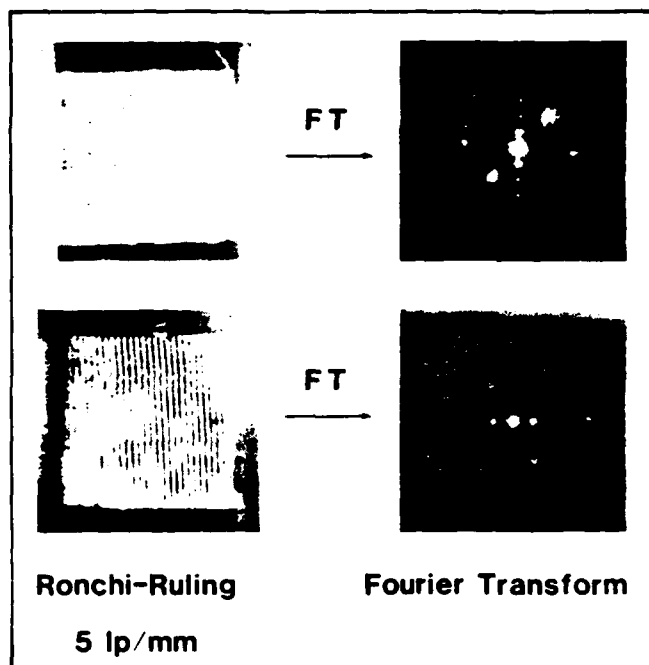


Fig. 8. Coherent replica images and corresponding Fourier transforms of a 5 line pairs/mm Ronchi ruling after photorefractive incoherent-to-coherent conversion. In the upper case K_G and K_S are perpendicular, whereas in the lower case K_G and K_S are parallel.

the crystal. As can be seen from the figure, the optimum focal point occurs in the volume of the crystal (in the center when in addition $\alpha_S d \ll 1$) and is not localized on the front surface of the crystal. The resolution will then be proportional to $(W/2)^{-1}$, which is in turn equal to $4nF\#/d$, where n is the refractive index of the electrooptic crystal, $F\#$ is the f-number of the input incoherent imaging system (assumed 1:1), d is the crystal thickness, and α_G and α_S are the absorption coefficients at the grating and signal wavelengths (λ_G and λ_S), respectively. For example, for $n = 2.5$, $d = 1$ mm, and an f-number of 5, the resolution limit is approximately 50 line pairs/mm. Figure 7 describes the case for which $\alpha_G d \gg 1$, such that the induced holographic grating has significant amplitude only within a thin layer of thickness $d_{eff} = \alpha_G^{-1}$. In this case, the resolution will be given by $4nF\#\alpha_G$. For $n = 2.5$, $d = 1$ mm, $\alpha_G = 100$ cm $^{-1}$, and an f-number of 5, the resolution limit is approximately 500 line pairs/mm. It should be noted that if the absorption coefficient α_S is chosen to be significantly larger than α_G , the resolution will be constrained by α_S instead of α_G .

Several systemic resolution constraints derive from the utilization of a volume hologram in the image conversion process. One such constraint is that the coherent grating spatial frequency must be significantly larger than the signal (incoherent image) bandwidth in order to effect separation of the diffracted beam components, due to harmonic generation that occurs simultaneously around the zeroth and first orders of the carrier grating. In addition, the range of spatial frequencies distributed about the coherent grating spatial frequency should be in an optimum region of the modulation transfer characteristic of the device.

A third fundamental systemic constraint on PICOC device resolution derives from Bragg detuning effects associated directly with the use of a volume hologram. This effect is strikingly illustrated in Fig. 8, which comprises converted images of a 5 line pair/mm Ronchi ruling and their related coherent Fourier transforms for two orthogonal orientations of the incoherent grating wave vector relative to the coherent grating wave vector. As can be seen from the figure, a significant difference in resolution exists between cases in which the

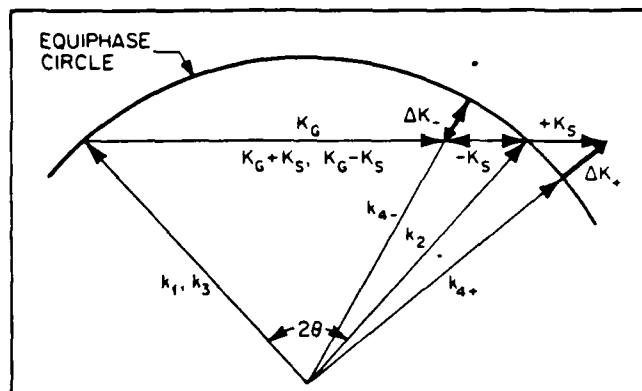


Fig. 9. Wave vector diagram of the PICOC process for the case of collinear signal (K_S) and grating (K_G) wave vectors. Readout with a beam incident at wave vector k_3 results in diffracted beams at k_{4+} and k_{4-} , with corresponding Bragg mismatches ΔK_+ and ΔK_- . The grating wave vector K_G is related to the wave vectors k_1 and k_2 of the coherent writing beams by $|K_G| = 2|k_1| \sin \theta = 2|k_2| \sin \theta$.

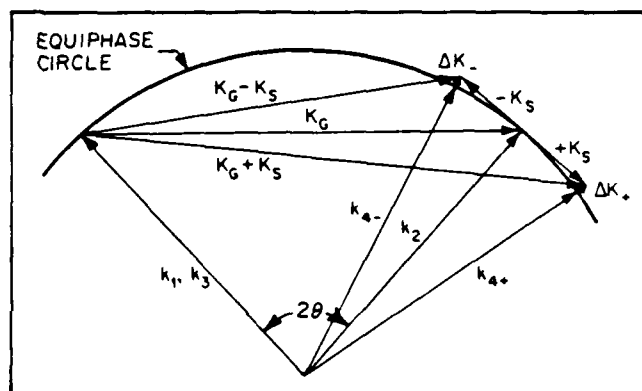


Fig. 10. Wave vector diagram of the PICOC process for the case in which the signal wave vector is tangent to the equiphase circle. Readout with a beam incident at wave vector k_3 results in diffracted beams at k_{4+} and k_{4-} , with significantly reduced Bragg mismatches ΔK_+ and ΔK_- .

wave vector of the ruling (incoherent grating) is parallel or perpendicular to the coherently written (holographic) grating. This difference derives principally from the fact that a different wave vector matching condition exists for these two cases, as shown in Figs. 9 and 10. In Fig. 9, the incoherent grating wave vector is parallel to the coherent grating wave vector, a condition achieved by symmetrically disposing the incident coherent beams about the normal to the crystal while simultaneously arranging the incoherent imaging system such that its optical axis is parallel to the crystal normal. In this case, significant Bragg detuning occurs for even small incoherent grating wave vectors (resulting in angular deviations of the diffracted beam). In Fig. 10, the incoherent grating wave vector is arranged to lie tangent to the circle defined by the incident coherent grating wave vectors, such that a significantly increased angular deviation of the diffracted beam is allowed before significant Bragg detuning effects occur. Such a wave vector tangency condition is automatically satisfied when the incoherent grating wave vector is normal to the coherent grating wave vector (as it is in the orientation normal to the plane of incidence). For the arrangement described in Fig. 10, wave vector tangency is assured for both orientations, so that the horizontal and vertical resolutions become degenerate. This is not the case for the situation depicted in Fig. 9, which explains the observations noted in Fig. 8. An equivalent condition is described in Ref. 7 for the case of distinct reading and writing beam wavelengths.

A fourth related systemic constraint involves the trade-off

between saturation diffraction efficiency and achievable resolution due to grating thickness. Within limits, increases in grating thickness increase the diffraction efficiency, but at the same time increase the Bragg detuning effects. This condition creates a familiar resolution/sensitivity trade-off situation in which one can be optimized at the expense of the other.

Additional resolution limitations derive from material-dependent parameter constraints that influence the physics of grating formation. This limitation is essentially described by the behavior of the diffraction efficiency as a function of the spatial frequency with the applied transverse electric field as a parameter.¹¹ The amplitude modulation transfer function of the grating itself depends on the spatial frequency dependence of the maximum space-charge field, which in turn depends on carrier mobilities and lifetimes, the applied field, the nature and distribution of traps in the crystal, and the trap occupancy (Fermi) level.

Optimization of the PICOC device resolution necessarily involves simultaneous satisfaction of all three types of constraints, and must be undertaken also in compromise with factors that directly affect the device sensitivity. In the following section, we present a model that allows the device sensitivity to be understood and optimized.

2.3. Theoretical model

The following simple model describes the encoding of the incoherent information onto the coherent diffracted beam by taking into account only the effective modulation ratio of the exposure in the bulk of the photorefractive material. The different intensities incident on the crystal are as shown in Fig. 2. The intensity pattern resulting from simultaneous illumination with the coherent writing beams and the incoherent image-bearing beam is given by Eq. (1), which describes the case of a one-dimensional incoherent sinusoidal grating with a fundamental component at wave vector K_S . The space-charge field E_{sc} is related to the current density $J(x,t)$ and the free-carrier concentration $n(x,t)$ through the following system of equations¹²:

Current density equation:

$$J(x,t) = eD \frac{\partial n(x,t)}{\partial x} + e\mu n(x,t) [E_{sc}(x,t) + E_0]; \quad (3)$$

Continuity equation:

$$\frac{\partial n(x,t)}{\partial t} = g(x) - \frac{n(x,t) - n_d}{\tau} + \frac{1}{e} \frac{\partial J(x,t)}{\partial x}; \quad (4)$$

Maxwell's equation (integral form):

$$E_{sc}(x,t) = -\frac{1}{\epsilon} \int_0^L J(x,t) dt + G(t); \quad (5)$$

Charge conservation equation (integral form):

$$\int_0^L E_{sc}(x,t) dx = 0. \quad (6)$$

in which $g(x)$ is the generation rate, n_d is the free-carrier concentration in the dark, E_0 is the applied transverse field, and $G(t)$ is a boundary-condition-dependent constant of integration. If one assumes that the quantum efficiency ξ , which is the probability of creating a free electron per unit absorbed photon of energy $\hbar\omega$, is not wavelength-dependent in the range of interest, the intensity distribution $I_T(x)$ gives rise to the free-carrier generation rate:

$$g(x) = \frac{\alpha_G \xi}{\hbar\omega_G} I_G (1 + m_G \cos K_G x) \exp(-\alpha_G d) + \frac{\alpha_S \xi}{\hbar\omega_S} I_S (1 + m_S \cos K_S x) \exp(-\alpha_S d), \quad (7)$$

in which d is the crystal thickness and the intensity absorption coefficients are α_S at λ_S and α_G at λ_G . Since free-carrier redistribution will occur both by drift and diffusion, the resultant space-charge field will depend on both the carrier concentration $n(x)$ and its spatial gradient $\partial n / \partial x$. Analytic solution of this system of equations is extremely complicated in general and must be accomplished in the context of boundary conditions applicable to a given experimental situation. In the development that follows, several simplifying assumptions are made that are appropriate for the simultaneous erasure/writing mode configuration in the steady state and in the large transverse applied field limit (characterized by enhanced saturation diffraction efficiency).

In the saturation regime (steady state), the current density is constant and the carrier concentration is given by

$$n(x) = n_d + \tau g(x), \quad (8)$$

in which the free-carrier lifetime τ is assumed to be the same for the writing and the erasure beams and to be unaffected by the presence or absence of illumination. In the strong drift regime for which the applied field is much larger than the diffusion field ($E_0 \gg E_D$; $E_D = Kk_B T/e$), solution of the system of equations [Eqs. (3) through (6)], subject to the additional constraints expressed in Eqs. (7) and (8), yields the following expression for the space-charge field:

$$E_{sc} = \frac{m_G^{eff} E_{DG} \sin K_G x + m_S^{eff} E_{DS} \sin K_S x}{1 + m_G^{eff} \cos K_G x + m_S^{eff} \cos K_S x} + E_0 \left\{ \frac{[1 - (m_G^{eff})^2]^{1/2}}{1 + m_G^{eff} \cos K_G x + m_S^{eff} \cos K_S x} - 1 \right\}, \quad (9)$$

with

$$m_G^{eff} = \frac{m_G}{1 + \frac{\alpha_S \lambda_S}{\alpha_G \lambda_G} \frac{I_S}{I_G} \exp[(\alpha_G - \alpha_S)d]}, \quad (10)$$

$$m_S^{eff} = \frac{m_S}{1 + \frac{\alpha_S \lambda_S}{\alpha_G \lambda_G} \frac{I_S}{I_G} \exp[(\alpha_G - \alpha_S)d]}. \quad (11)$$

Assuming further that the effective fringe modulation ratios (m_G^{eff} and m_S^{eff}) are small and that $E_{DS} \ll E_{DG} \ll E_0$, the space-charge field amplitudes at each harmonic component of interest are given by

$$E_{sc}(K_G) = -m_G^{eff} E_0, \quad (12)$$

$$E_{sc}(K_G + K_S) = m_G^{eff} m_S^{eff} E_0, \quad (13)$$

$$E_{sc}(K_G - K_S) = m_G^{eff} m_S^{eff} E_0. \quad (14)$$

The refractive index modulation induced along a principal axis by the electrooptic effect is then given by

$$\begin{aligned}\Delta n(x) &= \frac{1}{2} n_0^3 r_{41} E_{sc}(x) \\ &= -\frac{1}{2} n_0^3 r_{41} m_G^{eff} E_0 \{ \cos K_G x \\ &\quad - m_S^{eff} \cos[(K_G + K_S)x] - m_S^{eff} \cos[(K_G - K_S)x] \}. \quad (15)\end{aligned}$$

or, by combining terms,

$$\Delta n(x) = -\frac{1}{2} n_0^3 r_{41} m_G^{eff} E_0 (1 - 2m_S^{eff} \cos K_S x) \cos K_G x \quad (16)$$

For small diffraction efficiencies and low spatial frequencies K_S , the diffracted intensity pattern in the first order of the carrier grating can be expressed by the relation

$$I_4(x) = I_3 \left(\frac{\pi n_0^3 r_{41} m_G^{eff} E_0}{\lambda_G \cos \theta} \right)^2 (1 - 4m_S^{eff} \cos K_S x) \quad (17)$$

in the Kogelnik formulation,¹⁶ in which the angle θ is defined as shown in Fig. 2. This expression demonstrates that the information contained in the incoherent incident intensity distribution $I_S(x)$ [see Eq. (1)] is transferred onto the diffracted beam, performing an incoherent-to-coherent conversion in which the original contrast is reversed.

For the case of uniform erasure ($K_S = 0$), the diffracted beams I_{11} and I_{10} are degenerate (see Fig. 3). For the case of modulated erasure ($K_S \neq 0$), this degeneracy is broken and separate beams are observed. The diffraction efficiencies in these two cases are given by

$$\eta(I_{10}) \propto [m_G^{eff}(R)]^2 E_0^2 \quad (18)$$

$$\eta(I_{11}) \propto [m_G^{eff}(R) m_S^{eff}(R)]^2 E_0^2 \quad (19)$$

in which R is the ratio of the incoherent to the coherent beam intensities ($R = I_S/I_G$). Figures 11 and 12 illustrate the normalized diffraction efficiency of the I_{10} beam for three typical erasure wavelengths (λ_S). The parameter A_S is the ratio of the erasure beam to the writing beam intensity absorption coefficients ($A_S = \alpha_S/\alpha_G$). It is interesting to note that the maximum energy sensitivity does not generally occur for wavelength matching ($\lambda_G = \lambda_S$). In fact, two limiting situations can be recognized. For a thin crystal (Fig. 11), the optimum sensitivity is obtained with a strongly absorbed erasure beam because the generation rate increases with the ratio A_S , while the attenuation factor $\{\exp[(\alpha_G - \alpha_S)d]\}$ has almost no effect on the effective modulation m_G^{eff} . Conversely, for a thick crystal (Fig. 12), the exponential term dominates and the optimum sensitivity is achieved when the incoherent source has the same wavelength as the laser with which the carrier grating is recorded. This is intuitively appealing since the phase grating is most effectively erased throughout the volume of the crystal for the case of wavelength matching.

The normalized diffraction efficiency for the beam I_{11} in the case of modulated erasure is shown in Fig. 13. As predicted by Eq. (19), a maximum exists for $R_{eff} = 1$:

$$R_{eff} = R \frac{\alpha_S \lambda_S}{\alpha_G \lambda_G} \exp[(\alpha_G - \alpha_S)d] \quad (20)$$

The existence of such a maximum is explained as follows. It is clear that when $R_{eff} = 0$, there is no diffraction in the I_{11} direction. In addition, when R_{eff} is large, the uniform component of the incoherent exposure almost completely erases the carrier grating and the diffracted intensity vanishes. Therefore, the diffracted component in the (11) direction exhibits a maximum for some intermediate value of R_{eff} .

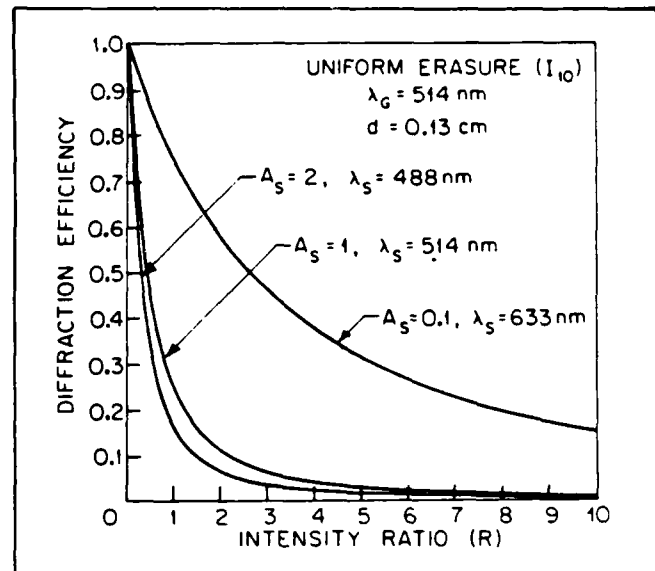


Fig. 11. Normalized diffraction efficiency of the I_{10} beam (uniform erasure) as a function of the intensity ratio (R) for a 0.13 cm thick crystal. Three signal beam (S) wavelengths (λ_S) with corresponding normalized absorption coefficients (A_S) are used to erase the coherent grating written at wavelength λ_G .

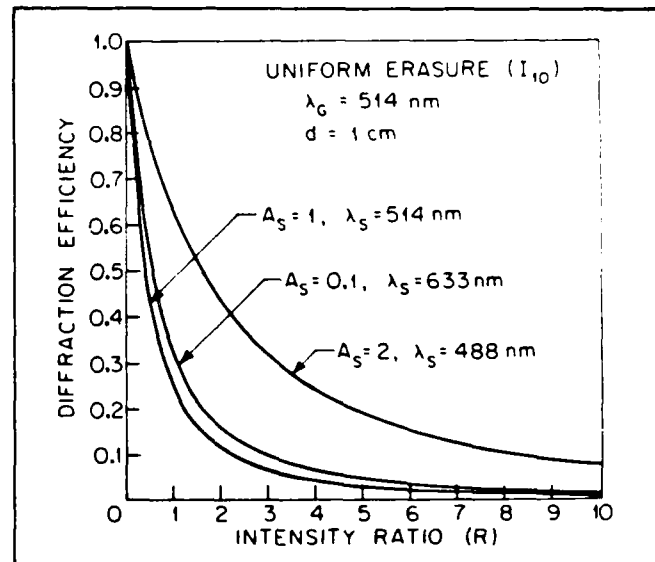


Fig. 12. Normalized diffraction efficiency of the I_{10} beam (uniform erasure) as a function of the intensity ratio (R) for a 1.0 cm thick crystal. Three signal beam (S) wavelengths (λ_S) with corresponding normalized absorption coefficients (A_S) are used to erase the coherent grating written at wavelength λ_G .

It should be noted that a relatively small intensity ratio ($R_{eff} \sim 1$) generates the maximum signal (I_{11}), but the uniform background (I_{10}) erasure is weak, resulting in low overall contrast. Conversely, a large ratio ($R_{eff} \gg 1$) decreases the modulation m_S^{eff} , but at the same time diffraction in the beam I_{10} is minimized so that the overall contrast is enhanced at the expense of sensitivity. In Figs. 4 and 5, R was set equal to 10 in order to maximize the contrast in the negative imaging mode. However, a number of specialized readout techniques (e.g., Schlieren and phase contrast) could be employed to simultaneously optimize the signal intensity and the contrast ratio.

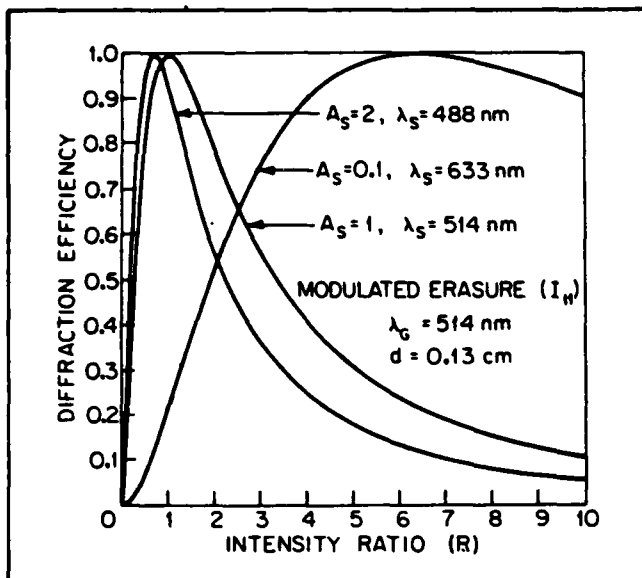


Fig. 13. Normalized diffraction efficiency of the I_{11} beam (modulated erasure) as a function of the intensity ratio (R) for a 0.13 cm thick crystal. Three signal beam (S) wavelengths (λ_S) with corresponding normalized absorption coefficients (A_S) are used to erase the coherent grating written at wavelength λ_G .

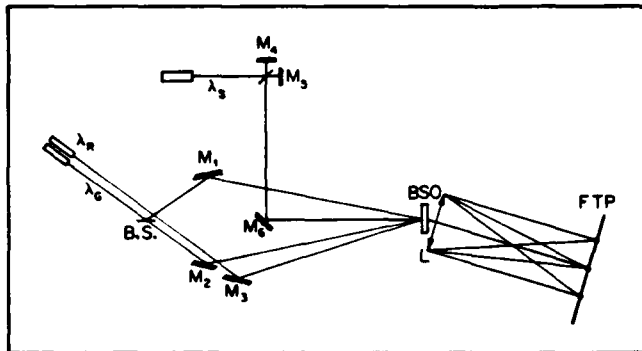


Fig. 14. Experimental arrangement for sensitivity and transfer function measurements, as described in detail in the text.

2.4. Experimental results

To verify the essential predictions of the model described above, the experimental arrangement shown in Fig. 14 was utilized. The carrier grating is recorded in the BSO crystal at $\lambda_G = 514 \text{ nm}$ and read out with a He-Ne laser at $\lambda_R = 633 \text{ nm}$ incident at the Bragg angle. The incoherent grating is formed with a Michelson interferometer (M_4, M_5) using the blue line of the argon laser ($\lambda_S = 488 \text{ nm}$). The far-field diffraction pattern is located in the Fourier plane of the lens L . The results of theoretical calculations and experimental measurements of the normalized diffraction efficiency as a function of the intensity ratio R are presented in Fig. 15 for the case of uniform erasure (I_{10}) and in Fig. 16 for the case of modulated erasure (I_{11}). The general behavior of these curves is in agreement with the model. The discrepancy in the amplitude between the theoretical prediction and the experimental measurement likely derives in part from a large difference in grating modulation ratio between the two treatments. Experimentally, a large modulation ratio ($m_G \sim 1$) was employed in order to increase the diffracted intensity for the purpose of signal-to-noise ratio enhancement. In the theoretical development, however, the assumption of small modulation ratios ($m_G \ll 1$) was necessary

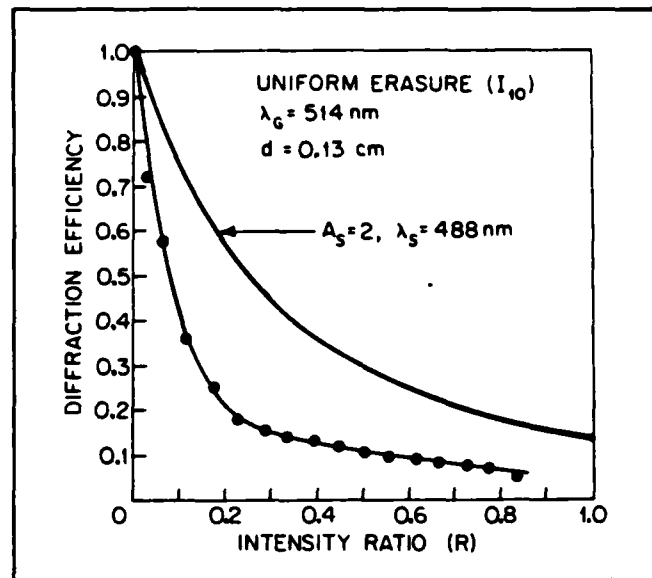


Fig. 15. Normalized experimental diffraction efficiency of the I_{10} beam (uniform erasure) as a function of the intensity ratio (R). The corresponding theoretical prediction is also shown for comparison.

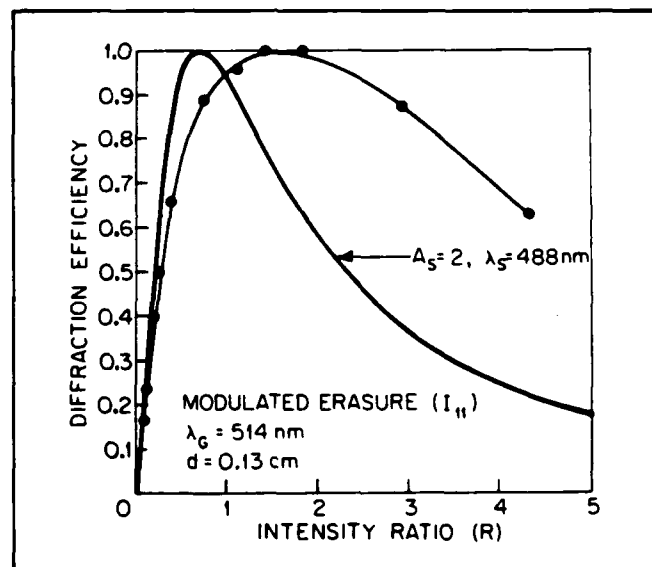


Fig. 16. Normalized experimental diffraction efficiency of the I_{11} beam (modulated erasure) as a function of the intensity ratio (R). The corresponding theoretical prediction is also shown for comparison.

to achieve analytic solution of the nonlinear coupled equations. An increase in the modulation ratio results in increased space-charge-field gradients that in turn enhance erasure sensitivity, as indicated by the experimental results shown in Fig. 15. This heightened sensitivity is strongly diminished at larger intensity ratios due to the reduced effective modulation ratio that results from uniform grating erasure. This explanation of the uniform erasure case also aids in understanding the behavior shown in Fig. 16 for the modulated case (I_{11}). At the larger (experimental) modulation ratio, the uniform component of the incoherent erasure beam is more effective in diminishing the carrier grating (represented by I_{10} —see Fig. 15) than the (low modulation ratio) theory predicts, resulting in a decrease in sensitivity of I_{11} at low values of R . On the other hand, for values of R

much greater than 1, the carrier grating is erased more slowly than in the theoretical prediction. This results in a more gradual decline in the I_{11} diffraction efficiency with increasing R than would occur at lower modulation ratios (see Fig. 16).

2.5. Response time considerations

The general temporal response behavior of the I_{11} (information-bearing) diffracted beam is shown in Fig. 17 for the two cases $R = 1.5$ and $R = 5.0$. In this experiment, the coherent grating is established in the saturation regime at time $t = 0$, at which point the incoherent erasure beam is allowed to expose the crystal. For a small ratio R ($I_S = 0.6 \text{ mW/cm}^2$ in Fig. 17), the diffraction efficiency increases steadily with time until it reaches a maximum. However, for a strong incoherent beam ($I_S = 2 \text{ mW/cm}^2$ in Fig. 17), a transient effect appears. Initially, the beam I_{10} is quite intense and diffracts from the composite grating at wave vector $K_G + K_S$ to generate a rapid rise in the amplitude of I_{11} , but the uniform part of the incoherent illumination simultaneously erases the coherent carrier grating and hence the diffraction efficiency decreases to a small steady-state value. The time constant for this particular set of experimental parameters is in the range 0.5 to 1.5 s, although the device response time can be decreased by increasing the intensity of the carrier grating writing beams. In addition, the response time is strongly dependent on a number of intrinsic and extrinsic material parameters. The full range of variation possible for the temporal behavior characteristic is not firmly established at this point.

3. CONCLUSION

We have demonstrated the feasibility of real-time incoherent-to-coherent conversion utilizing phase conjugation in photorefractive BSO crystals. The physical basis of this effect is explained with a simple model that takes into account the influence of the effective modulation ratio on the diffraction efficiency of the different diffracted beams. Although the results are preliminary, the device proposed in this paper has potential for incoherent-to-coherent conversion with high resolution, which can be realized by optimizing the wave vector matching condition and the relative intensities and wavelengths of various beams. In addition, the PICOC device is quite attractive from considerations of low cost, ease of fabrication, and broad availability. With such a device, numerous optical processing functions can be directly implemented that utilize the flexibility afforded by the simultaneous availability of incoherent-to-coherent conversion and volume holographic storage.

4. ACKNOWLEDGMENTS

The authors thank F. Lum, D. Seery, M. Garrett, and Y. Shi for their technical assistance. This research was supported in part at the University of Southern California by the Air Force Systems Command (RADC) under Contract No. F19628-83-C-0031, the Defense Advanced Research Projects Agency (Office of Naval Research), the

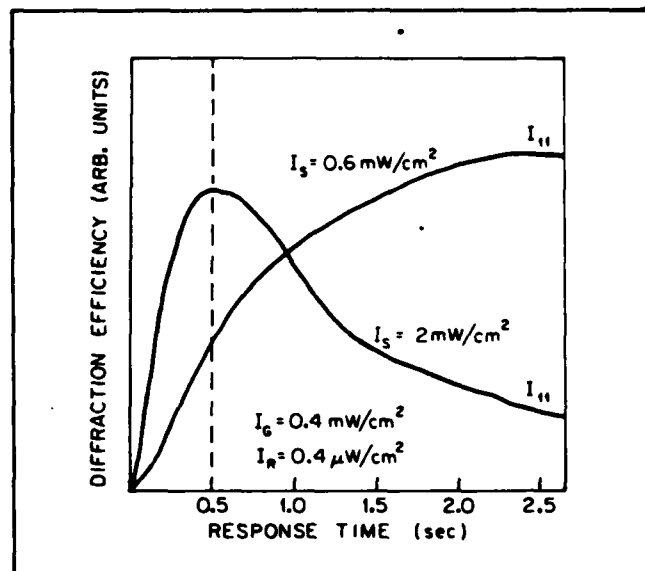


Fig. 17. Measured diffraction efficiency of the I_{11} beam (modulated erasure) as a function of time for two values of the signal intensity.

Joint Services Electronics Program, and the Army Research Office; and at the California Institute of Technology by the Air Force Office of Scientific Research and the Army Research Office.

5. REFERENCES

1. J. O. White and A. Yariv, *Appl. Phys. Lett.* 37, 5 (1980).
2. A. Marrakchi, J. P. Huignard, and J. P. Herriau, *Opt. Commun.* 34, 15 (1980).
3. G. J. Dunning and R. C. Lind, *Opt. Lett.* 7, 558 (1982).
4. A. R. Tanguay, Jr., *Proc. Army Research Office Workshop on Future Directions for Optical Information Processing*, J. Walkup, ed., pp. 52-77 (1981).
5. D. Casasent, *Proc. IEEE* 65, 143 (1977).
6. Y. Shi, D. Psaltis, A. Marrakchi, and A. R. Tanguay, Jr., *Appl. Opt.* 22, 3665 (1983).
7. A. A. Kamshilin and M. P. Petrov, *Sov. Tech. Phys. Lett.* 6, 144 (1980).
8. D. L. Staebler and J. J. Amodei, *J. Appl. Phys.* 43, 1042 (1972).
9. M. Peltier and F. Micheron, *J. Appl. Phys.* 48, 3683 (1977).
10. R. Grousson and S. Mallick, *Appl. Opt.* 19, 1762 (1980).
11. J. P. Huignard, J. P. Herriau, G. Rivet, and P. Gunter, *Opt. Lett.* 5, 102 (1980).
12. J. P. Herriau, J. P. Huignard, and P. Aubourg, *Appl. Opt.* 17, 1851 (1978).
13. Y. Owechko and A. R. Tanguay, Jr., *Opt. Lett.* 7, 587 (1982).
14. P. Aubourg, J. P. Huignard, M. S'areng, and R. A. Mullen, *Appl. Opt.* 21, 3706 (1982).
15. M. G. Moharam, T. K. Gaylord, and R. Magnusson, *J. Appl. Phys.* 50, 5642 (1979).
16. H. Kogelnik, *Bell Syst. Tech. J.* 18, 2909 (1969).

PHOTOREFRACTIVE INCOHERENT-TO-COHERENT

OPTICAL CONVERSION

D. Psaltis and J. Yu
Department of Electrical Engineering
California Institute of Technology
Pasadena, California 91125

and

A. Marrakchi and A. R. Tanguay, Jr.
Optical Materials and Devices Laboratory
Departments of Electrical Engineering and Materials Science
University of Southern California, Los Angeles, California 90089-0483

ABSTRACT

A method for performing incoherent-to-coherent conversion in photorefractive crystals is presented. The technique is experimentally demonstrated in $\text{Bi}_{12}\text{SiO}_{20}$ and a theoretical framework is established to analyze the performance of the device.

INTRODUCTION

Two dimensional optical transparencies are used in the implementation of optical image processing systems as input incoherent-to-coherent transducers, and are also used for the recording of Fourier transform holograms. Most of the early optical image processors were implemented using photographic film for both purposes. The need for real time operation has prompted the development of devices with which it is possible to form reusable optical transparencies at high speed [1, 2]. In general, however, real time devices that are intended for use as input spatial light modulators (SLMs) are not suitable for holographic recording and vice-versa. This distinction exists partially due to the fact that a much higher spatial bandwidth is necessary for holographic recording, while at the same time the properties of high resolution holographic recording media (low sensitivity, response at low spatial frequencies) are not always compatible with the requirements of an efficient incoherent-to-coherent conversion process. Such a resolution-sensitivity tradeoff is familiar from the example of photographic (silver halide) films, for which high spatial bandwidth is in general accomplished at the expense of lower sensitivity.

The dichotomy between input SLMs and holographic elements is particularly evident in devices that are fabricated with photorefractive (PR) crystals. Photorefractive crystals are photosensitive as well as electrooptic. This feature makes it possible to record an optical image in such crystals, and subsequently modulate a separate optical beam with the recorded image. Real time SLMs, such as the PROM [3], have been fabricated with the photorefractive material $\text{Bi}_{12}\text{SiO}_{20}$, and this same crystal has also been used extensively in real-time holographic experiments [4]. The configuration and the properties of the holographic devices, however, are quite different from the PROM and other similar SLMs that are fabricated with $\text{Bi}_{12}\text{SiO}_{20}$.

In this paper we examine a method for performing incoherent-to-coherent conversion with a photorefractive device in the holographic configuration. There are two primary reasons that have motivated us to investigate this possibility: first, the construction of the holographic device is inexpensive and simple, and second, it is possible to record very high spatial frequencies in photorefractive crystals in the holographic configuration. This technique has been recently demonstrated experimentally by Kamshilin and Petrov [5] and independently by the authors [6]. In this paper the method is described, experimental results are presented, and an analytical model is employed to determine the electric fields induced in the photorefractive crystal by this recording mechanism.

INCOHERENT-TO-COHERENT CONVERSION

The photorefractive incoherent-to-coherent optical converter is shown in Fig. 1. The photorefractive crystal is exposed to the interference pattern of two coherent plane waves, which induces the formation of a phase grating in the crystal. An electric field is externally applied in a direction parallel to the grating wavevector. The grating formation is attributed to photo-induced generation and subsequent retrapping of free carriers, resulting in a stored periodic space-charge field. The space-charge field introduces a proportional modulation of the index of refraction in the crystal through the electrooptic effect. In the configuration of Fig. 1, the crystal is also exposed to the incoherent image of an input object. The exposure to the incoherent intensity can take place during, after, or before the formation of the holographic grating. It is assumed that the depth of focus of the imaging system is sufficiently long so that the incoherent image is focused throughout the volume of the photorefractive crystal.

The incoherent illumination induces photogenerated carriers in the conduction band. As these carriers diffuse or drift and are retrapped, they partially cancel the holographically induced space-charge distribution, resulting in local erasure of the holographic grating in proportion to the incoherent intensity. The space-charge field that results from the combined exposure to the coherent and incoherent beams is a periodic grating whose amplitude is modulated approximately proportional to the negative of the incoherent image. A coherent reproduction of the incoherent image can be obtained by illuminating the crystal with an auxiliary laser beam at the appropriate Bragg angle and subsequently forming an image of the diffracted light.

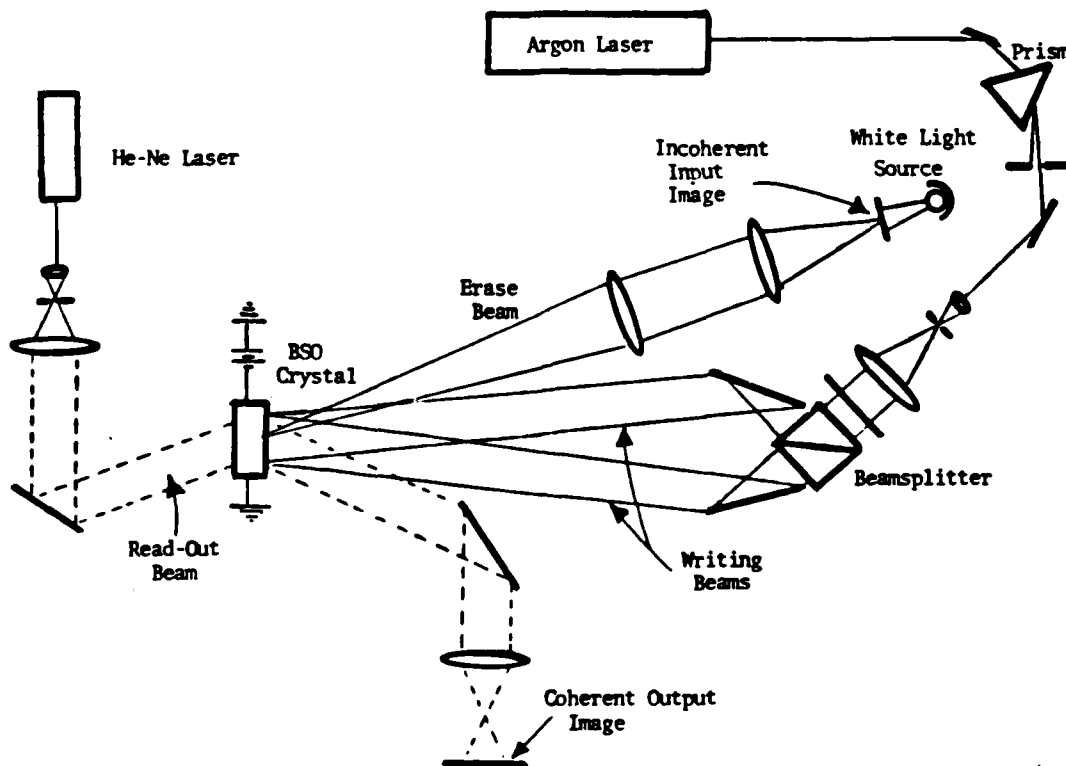


Fig. 1. Optical system for the implementation of incoherent-to-coherent conversion with a photorefractive crystal.

The system of Fig. 1 was assembled in the laboratory to demonstrate experimentally the feasibility of this technique. A (110) plate of single crystal $\text{Bi}_{12}\text{SiO}_{20}$ (BSO) with dimensions 10mm x 10mm x 2mm was used and an electric field equal to 6kV/cm was applied in the <001> direction. The holographic grating was formed with green light ($\lambda = 514 \text{ nm}$) derived from an Argon ion laser at a spatial frequency equal to 10 line pairs/mm. The average intensity of the coherent beam was 33 mW/cm^2 . The $\text{Bi}_{12}\text{SiO}_{20}$ crystal was exposed simultaneously to an incoherent image with average intensity 1.7 mW/cm^2 . The recorded image was read out with an expanded beam from a HeNe laser as shown in Fig. 4, and the coherent replica was obtained by imaging the plane of the $\text{Bi}_{12}\text{SiO}_{20}$ plate onto the output plane. Examples of images that were recorded after incoherent-to-coherent conversion are shown in Fig. 2.

A complete description of the performance that was observed in these initial experiments is presented in a separate paper in this volume [7]. We will briefly summarize some of the results here. Spatial frequencies in excess of 10 line pairs/mm were recorded and reconstructed, resulting in converted images with over 10 resolvable pixels with the 1 cm crystal used in the experiments. There is a direct trade-off between sensitivity and speed. Response times approximately equal to 30 msec were observed with relatively high incoherent intensity; at lower intensities ($\sim 1 \text{ mW/cm}^2$) the response time was approximately equal to 500 msec. It should be emphasized that these characteristics are simply preliminary observations at this stage. All of the performance characteristics depend strongly on materials and configuration parameters, and numerous trade-offs exist. The physical mechanisms that impose the fundamental limitations must be identified before the parameters can be chosen to yield optimum overall performance. To this end, a mathematical model is employed to analyze incoherent-to-coherent conversion in photorefractive crystals, as described in the following section.

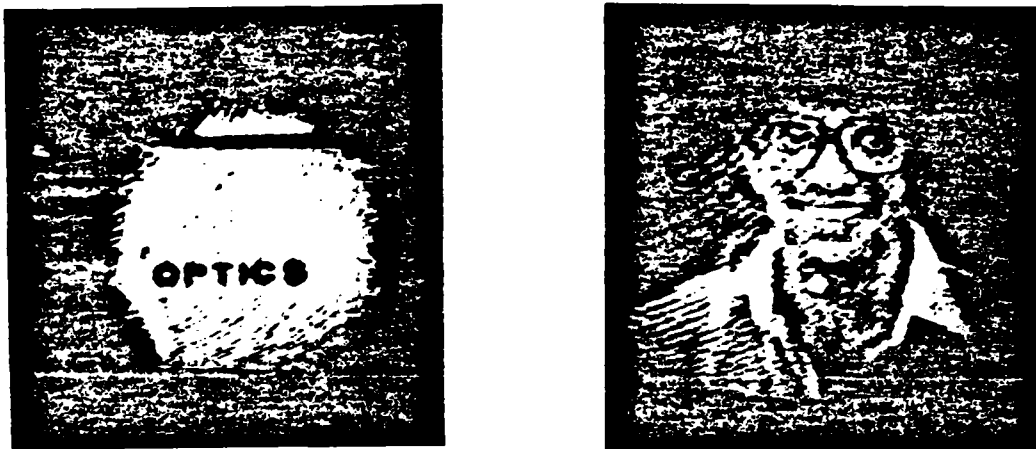


Fig. 2. Examples of photorefractive incoherent-to-coherent conversion.

MATHEMATICAL MODEL

Holographic grating formation in photorefractive crystals has been extensively modeled [8]. The same basic formalism can also be used to analyze incoherent image recording in photorefractive crystals through erasure of a holographic grating. The analytical framework that has been proposed by Kukhtarev [8] is based on the following set of equations:

$$\frac{\partial n}{\partial t} = \frac{\partial N_D^i}{\partial t} + \frac{1}{e} \frac{\partial J}{\partial x} \quad (1)$$

$$\frac{\partial N_D^i}{\partial t} = sI(N_D - N_D^i) - \gamma n N_D^i \quad (2)$$

$$\frac{\partial E}{\partial x} = \frac{e}{\epsilon} (N_D^i - n - N_0) \quad (3)$$

$$J = en\mu E + k_B T \mu \frac{\partial n}{\partial x} \quad (4)$$

$$\nabla^2 \vec{E} + \omega^2 \mu E (1 + n^2 r E) \vec{E} = 0 \quad (5)$$

in which

n	is the density of electrons in the conduction band
N_D^i	is the density of vacant donor sites
e	is the electron charge
J	is the current density
x	is the coordinate along which the external voltage is applied
s	is the absorption cross section
I	is the intensity of the incident optical wave
N_D	is the total density of donor sites
γ	is the recombination coefficient
E	is the electric field in the crystal
ϵ	is the dielectric constant
N_0	is the density of vacant acceptor sites in the dark
μ	is the mobility
k_B	is Boltzmann's constant
T	is the temperature
E	is the amplitude of the electric field of the read-out light
$\omega/2\pi$	is the optical frequency
r	is the electrooptic coefficient
n	is the index of refraction

Equations (1)-(4) can be solved to determine the electric field E that is induced in the crystal by exposure to the input optical intensity I . Equation (5) is the wave equation, including the polarization field that is induced by the electric field in the crystal through the electrooptic effect. Once the electric field E is determined from Eqs. (1)-(4), Eq. (5) can be solved using coupled mode analysis [10] to determine the optical field diffracted by the photorefractive crystal. In this paper we concentrate on determining only the induced field E . We consider the simultaneous exposure of the crystal to the coherent grating and an incoherent image. The total light intensity incident on the crystal is the sum of the coherent and incoherent intensities:

$$I(x, t) = I_c [1 + m_c \cos(K_c x)] + I_i(x, t) \quad (6)$$

where I_c is the average intensity of the coherent light, m_c is the modulation depth and $K_c/2\pi$ is the spatial frequency of the coherent grating. The intensity of the incoherent image is given by $I_i(x, t)$.

The set of Eqs. (1)-(4) describe a system with input intensity $I(x, t)$, the state of which is described by the four variables N_D , n , J and E . The solution of the four equations yields the electric field E . The fidelity with which the incoherent image I_i is recorded

can be investigated by studying the relationship between E and I_i . The solution of Eqs. (1)-(4) is in general difficult because they are nonlinear. Often these equations are solved by making simplifying assumptions that lead to full or partial linearization. The effect that we are investigating, however, is based on the nonlinearities, and we are therefore primarily interested in characterizing the nonlinear part of the response of the system. For this reason, we have included in the analysis all of the nonlinear terms. The strongest nonlinearity typically occurs in Eq. (4) where the current density J is proportional to the product of E and n , the number density of electrons in the conduction band. The product nN_D that appears in Eq. (1) can also be a significant source of nonlinearity.

The critical role that the nonlinearities play in this device can be appreciated by considering the electric field distributions E that are induced by exposure to the coherent and incoherent beams independently. The coherent beams alone record a uniform grating; the incoherent beam alone induces a field distribution that is not an accurate reproduction of the incoherent intensity distribution (more on this later). If the system were linear, the combined exposure would simply induce the sum of these two field distributions. However, the nonlinearities give rise to additional, intermodulation terms in E and these are responsible for the accurate reproduction of the incoherent image upon reconstruction of the holographic grating that is observed experimentally.

The steady state behavior of the system is analyzed by setting all the time derivatives equal to zero in Eqs. (1)-(4) and choosing an incoherent illumination of the following form:

$$I_i(x, t) = I_i [1 + m_i \cos(K_i x)]. \quad (7)$$

in which m_i is the modulation depth and $K_i/2\pi$ the spatial frequency of the incoherent grating. In this case the solution for the electric field E is in the form

$$E(x) = \sum_{m=-M}^M \sum_{n=-N}^N E_{mn} \exp[j(mK_c + nK_i)x], \quad (8)$$

and similarly for the other three variables, n , N_D , and J . Equations (1)-(4) can now be solved numerically, using the method of successive approximations, to obtain the coefficients of the expansion in Eq. (8). The calculation was performed for $m_c = m_i = 1$, $K_c/2\pi = 10$ c/mm, $K_i/2\pi = 10$ c/mm, and material parameters that were given by Peltier and Micheron [9] for $\text{Bi}_{12}\text{SiO}_{20}$.

The coefficients E_{11} , E_{01} and E_{10} are plotted in Fig. 3 as a function of the ratio $R = I_i/I_c$. The quantity E_{11} is the coefficient of the intermodulation term of interest, $\cos(K_i x) \cos(K_c x)$. It is evident in Fig. 3 that E_{11} attains a maximum at $R = 1$. It is interesting to note that in the experiments reported in the previous section, the much higher value of $R \sim 52$ was used. The reason for this choice can be appreciated by observing the behavior of E_{10} . The quantity E_{10} is the coefficient of the $\cos(K_i x)$ term, and upon reconstruction it results in a uniform plane wave diffracted in the same general direction as the signal beam that is diffracted due to the E_{11} grating. As a result it acts as a constant background bias that reduces the contrast of the coherent image. As shown in Fig. 3, the ratio E_{11}/E_{10} increases (and hence the contrast improves) as R is increased. In a coherent optical system, a constant bias can be easily removed by spatial filtering and

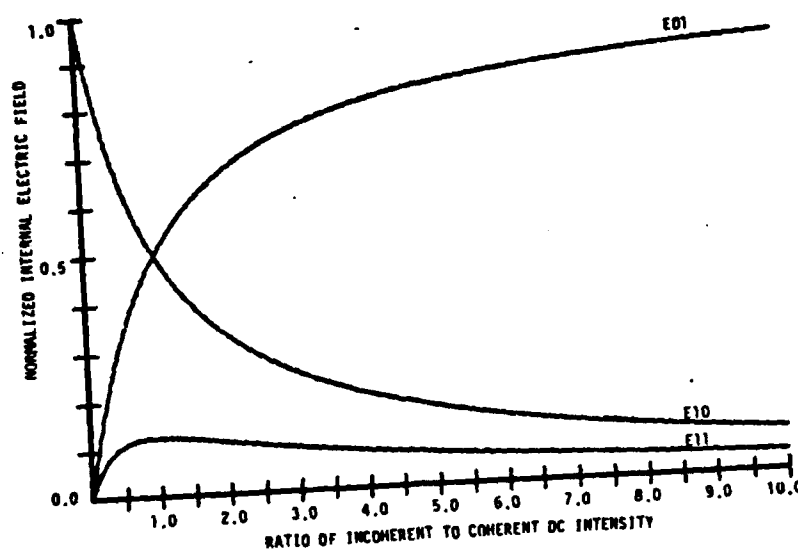


Fig. 3. Peak amplitude of three Fourier components of the electric field induced in the photorefractive crystal, as a function of the ratio of incoherent-to-coherent intensity.

therefore it is not necessary to use high incoherent intensity to obtain good contrast. A very significant conclusion that can be drawn is that the sensitivity for incoherent recording in this mode is equal to the sensitivity for holographic recording. Furthermore, a better approximation to linear incoherent recording is obtained in the range $0 \leq R \leq 1$ than at higher values of R .

The E_{01} coefficient is also plotted in Fig. 3. This term is due to the direct recording of the incoherent grating in the photorefractive crystal. As can be seen from Fig. 3, E_{01} is stronger than E_{11} and thus the question may arise as to whether or not direct incoherent image recording in photorefractive crystals is feasible in the holographic configuration. Unfortunately direct recording is only possible for one dimensional

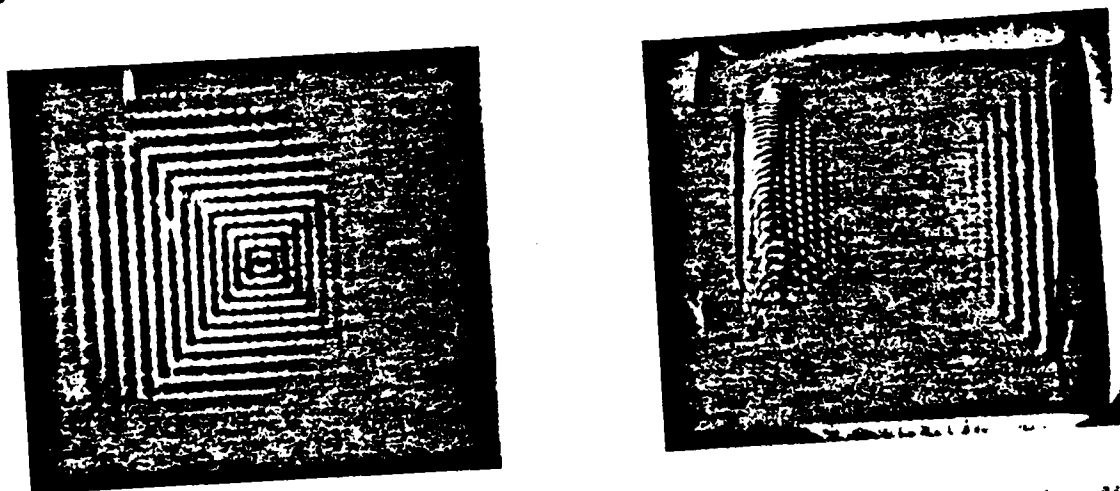


Fig. 4. Optical reconstruction of a nested squares pattern using (a) direct recording, and (b) holographic grating erasure techniques, as described in the text.

incoherent images in crystals such as $\text{Bi}_{12}\text{SiO}_{20}$, where an external field is required for efficient grating recording. A simple experiment was performed to demonstrate this. A transparency of a nested-squares pattern was used as the incoherent input in the experimental set-up described in the previous section. Coherent reproductions of this object were obtained from the E_{11} and E_{01} gratings and they are shown in Figs. 4a and 4b, respectively. The external field was applied in the horizontal direction of Figs. 4. It is evident that information is recorded only in the direction parallel to the external field with direct exposure, whereas two dimensional recording is possible through erasure of the holographic grating.

CONCLUSION

A novel recording method in photorefractive crystals has been presented. Our preliminary experimental and theoretical observations indicate that incoherent-to-coherent conversion can be implemented with this mechanism. Further experimental and theoretical work is needed to assess the full capabilities of this device.

ACKNOWLEDGMENTS

The research reported in this paper is supported at Caltech by the Army Research Office and the Air Force Office of Scientific Research; at USC by the Air Force Systems Command (RADC) under contract F19628-83-C-0031, the Defense Advanced Research Projects Agency, the Joint Services Electronics Program, and the Army Research Office.

The authors would like to thank Y. Shi for performing some of the experiments described in this paper.

REFERENCES

- [1] A. R. Tanguay, Jr., in Proceedings of ARO Workshop on Future Directions for Optical Information Processing, Texas Tech University, Lubbock, Texas, 52-77, (1981).
- [2] D. Casasent, Proc. IEEE., **67**(5), 813 (1979).
- [3] B. A. Horwitz and F. J. Corbett, Opt. Eng., **17**, 353, (1978).
- [4] R. A. Fisher, ed., Optical Phase Conjugation, Academic Press, New York, (1983).
- [5] A. A. Kamshilin and M. P. Petrov, Sov. Tech. Phys. Lett., **6**(3), 144, (1980).
- [6] Y. Shi, D. Psaltis, A. Marrakchi and A. R. Tanguay, Jr., Appl. Opt., **22**, 3665 (1983).
- [7] A. Marrakchi, A. R. Tanguay, Jr., D. Psaltis and J. Yu, this volume.
- [8] N. V. Kukhtarev, et. al., Ferroelec., **22**, 2414, (1975).
- [9] M. Peltier and P. Micheron, J. Appl. Phys., **45**, 3683, (1977).
- [10] H. Kogelnik, Bell Syst. Tech., **48**, 2909, (1969).

PHOTOREFRACTIVE INCOHERENT-TO-COHERENT OPTICAL CONVERTER:

PHYSICAL AND MATERIALS CONSIDERATIONS

A. Marrakchi and A. R. Tanguay, Jr.
Optical Materials and Devices Laboratory
Departments of Electrical Engineering and Materials Science
University of Southern California, Los Angeles, California 90089-0483

and

J. Yu and D. Psaltis
Department of Electrical Engineering
California Institute of Technology, Pasadena, California 91125

ABSTRACT

Volume holographic storage in photorefractive $\text{Bi}_{12}\text{SiO}_{20}$ (BSO) crystals is utilized to perform dynamic incoherent-to-coherent image conversion by means of selective spatial erasure of a uniform grating with white (or quasi-monochromatic) light. Preliminary results with binary and grey level transparencies are presented, and the conversion process is described in terms of a simple model which relates the diffracted intensity to the space-variant effective modulation ratio.

INTRODUCTION

Real-time holography in photorefractive materials has been utilized for many applications in the areas of optical processing [1], non-destructive testing of vibrating structures [2], and image propagation through aberrating media [3]. In addition, the development of high performance spatial light modulators has received much recent attention, primarily for incoherent-to-coherent conversion functions in optical information processing and computing applications [4, 5]. We have reported in a previous paper [6] the successful application of volume phase holography to the real-time conversion of an incoherent image (quasi-monochromatic or white light illumination) into a coherent replica. Several possible configurations for the performance of this image transduction in photorefractive materials were discussed therein. Such a conversion process has been investigated independently, as reported in reference [7].

In this paper, we present a simple model that describes the incoherent-to-coherent conversion process in terms of the space-variant effective modulation ratio. The model is shown to satisfactorily explain the principal features of the device sensitometry behavior. The limiting resolution of the process is discussed, and geometric, systemic, and materials-related constraints are identified. Experimental measurements of the device temporal response characteristics are presented.

The high sensitivity of photoconductive and electrooptic crystals such as Bismuth Silicon Oxide ($\text{Bi}_{12}\text{SiO}_{20}$ or BSO) in the visible spectrum has allowed the simultaneous recording and reading of volume holograms to be achieved with time constants amenable to real-time operation. The holographic recording process in photorefractive materials involves photoexcitation, charge transport, and trapping mechanisms [8, 9]. When two coherent writing beams are allowed to interfere within the volume of such a crystal, free carriers are nonuniformly generated by absorption, and are redistributed by diffusion and/or drift under the influence of an externally applied electric field. Subsequent trapping of these charges generates a stored space-charge field, which in turn modulates the refractive index through the linear electrooptic (Pockels) effect and thus records a volume phase hologram. If both coherent writing beams are plane waves, the induced hologram will consist of a uniform grating.

In the Photorefractive Incoherent-to-Coherent Optical Conversion (PICOC) process, in addition to the holographically-induced charge distribution stored in the crystal, an incoherent image is focused in the volume of the photorefractive material, and creates a spatial modulation that can be transferred onto a coherent beam by reconstructing the holographic grating. The spatial modulation of the coherent reconstructed beam will then be a negative replica of the input incoherent image. The holographic grating can be recorded before, during or after the crystal is exposed to the incoherent image. Therefore, a number of distinct operating modes are possible. These include the Grating Erasure Mode (GEM), the Grating Inhibition Mode (GIM), and the Simultaneous Erasure/Writing Mode (SEWM), among others.

In the Grating Erasure Mode, a uniform grating is first recorded by interfering the two coherent writing beams in the photorefractive crystal. This grating is then selectively erased by incoherent illumination of the crystal with an image-bearing beam. The incoherent image may be incident either on the same face of the crystal as the writing beams, or on the opposite face. When the absorption coefficients of the writing and image-bearing beams give rise to significant depth nonuniformity within the crystal, these two cases will have distinct wavelength-matching conditions for response optimization.

In the Grating Inhibition Mode, the crystal is pre-illuminated with the incoherent image-bearing beam prior to grating formation. This serves to selectively decay (enhance) the applied transverse electric field in exposed (unexposed) regions of the crystal. After this pre-exposure, the coherent writing beams are then allowed to interfere within the crystal, causing grating formation with spatially varying efficiency due to significant differences in the local effective applied field.

In the Simultaneous Erasure/Writing Mode, the incoherent image modulation, the coherent grating formation process, and the readout function are performed simultaneously. In one possible implementation of this mode, the conventional degenerate four-wave mixing geometry can be modified to include simultaneous exposure by an incoherent image-bearing illumination, as shown schematically in Fig. 1. In this configuration, the coherent replica of the incoherently-exposed image appears in the conjugate diffracted beam.

The PICOC process can thus be regarded as caused by selective spatial modulation of a grating by spatial encoding of an incoherent erasure beam. It should be noted here that a related image encoding process could be implemented in a non-holographic manner by pre-multiplication of the image with a grating [10].

PHOTOREFRACTIVE INCOHERENT-TO-COHERENT OPTICAL CONVERSION

In the remainder of this paper, we examine the Simultaneous Erasure/Writing Mode in more detail. As a specific example, we will discuss the configuration in which the coherent writing beams and the incoherent image-bearing beam illuminate the same face of a BSO photorefractive crystal, with a transverse electric field applied along the $\langle 110 \rangle$ axis as shown in Fig. 2. Consider an incident intensity distribution of the form

$$\begin{aligned} I_T(x) &= I_G(x) + I_S(x) \\ &= I_G(1 + m_G \cos K_G x) + I_S(1 + m_S \cos K_S x) \end{aligned} \quad (1)$$

with $I_G = I_1 + I_2$ and $m_G = 2(I_1 I_2)^{1/2} / I_G$, in which K_G and K_S are the wavevectors of the coherent grating (G) and a particular harmonic component of the incoherent image (S), and m_G and m_S are the respective modulation ratios. Due to the nonlinear dependence of the induced space-charge field on the incident intensity, such a distribution will generate a refractive index modulation which can be written in the form

$$\begin{aligned} \Delta n(x) &= C_1 \cos K_G x + C_2 \cos K_S x \\ &+ C_3 \cos [(K_G - K_S)x] \\ &+ C_4 \cos [(K_G + K_S)x] \\ &+ \text{Other Terms.} \end{aligned} \quad (2)$$

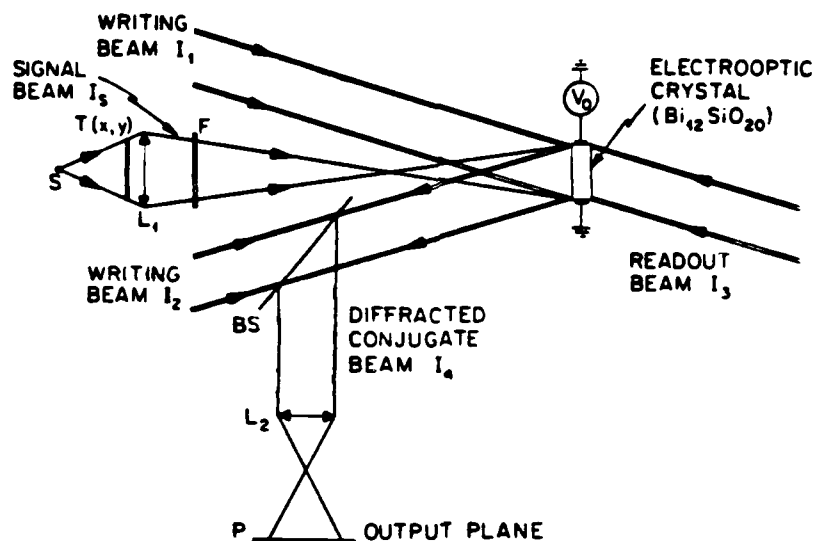


Fig. 1 Experimental configuration for photorefractive incoherent-to-coherent optical conversion (PICOC) in the Simultaneous Erasure/Writing Mode (SEWM). The writing beams I_1 and I_2 , and the reading beam I_3 , are generated from an argon laser ($\lambda = 514 \text{ nm}$). The phase conjugate beam I_4 is diffracted at the same wavelength. The transparency $T(x, y)$ is illuminated with a xenon arc lamp S and imaged onto the $\text{Bi}_{12}\text{SiO}_{20}$ crystal with the optical system L_1 through a filter $F(\lambda = 545 \text{ nm})$. A coherent replica is formed in the output plane by utilizing beamsplitter BS in conjunction with lens L_2 to image the surface of the crystal onto the output plane through polarizer P.

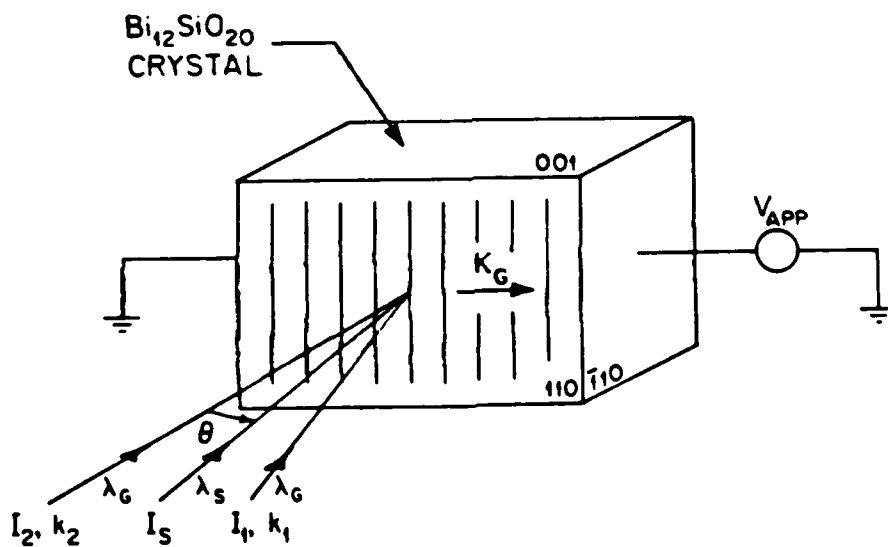


Fig. 2 PICOC transverse electrooptic configuration and recording geometry. The volume holographic grating with wavevector K_G is formed by the coherent writing beams I_1 and I_2 , and the incoherent information is encoded on the beam I_S .

Note in particular that $n(x)$ contains the sum and difference spatial frequencies as well as the fundamental harmonics. The constants C_i ($i = 1$ to 4) represent the amplitudes of each frequency component. Thus, the far-field diffraction pattern shown in Fig. 3 will comprise several beams, each corresponding to a specific term in the refractive index modulation.

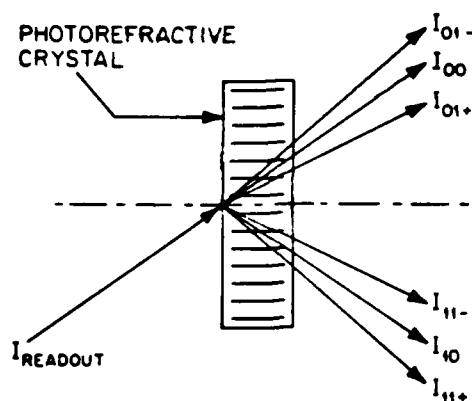


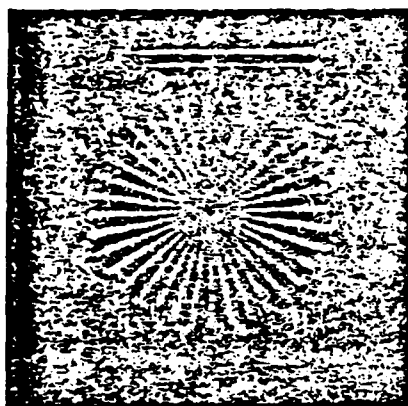
Fig. 3 Far-field diffraction pattern resulting from the nonlinear interaction of two gratings in a photorefractive crystal.

To demonstrate incoherent-to-coherent conversion in the four-wave mixing configuration, the experimental arrangement of Fig. 1 was utilized with a 1.3 mm thick (110)-cut BSO crystal. A transverse electric field ($E_0 = 4$ kV/cm) was applied along the $\langle 110 \rangle$ axis, and the carrier frequency of the holographic grating ($f = 300$ line pairs/mm) was chosen to lie within the optimum range for drift-aided charge transport in BSO [11].

In this implementation, the two plane wave writing beams (labeled I_1 and I_2) are generated from an argon laser ($\lambda_G = 514$ nm) and interfere inside the (110)-cut BSO crystal to create a volume phase hologram. The readout beam I_3 , collinear with I_1 to satisfy the Bragg condition, is diffracted to form the phase conjugate beam I_4 at the same wavelength. Increased diffraction efficiency results when a transverse electric field is applied to the electrooptic medium. An incoherently illuminated transparency $T(x, y)$, with intensity $I_S(x, y)$ (either quasi-monochromatic or white light), is imaged in the plane of the crystal. The beam-splitter BS separates the diffracted signal from the writing beam, and the polarizing filter P in the output plane eliminates the unwanted scattered light to enhance the signal-to-noise ratio [12]. The vertically polarized coherent writing beam and signal intensities were $I_{1,2} = 0.4$ mW/cm² and $I_S = 8$ mW/cm², respectively. The incoherent illumination was provided by a xenon arc lamp through a broadband filter centered at $\lambda_S = 545$ nm (FWHM = 100 nm).

The converted images obtained from two binary transparencies (a spoke target and a USAF resolution target), and from a black and white slide with grey levels are shown in Figs. 4 and 5, respectively. The original transparency and its converted image have reversed contrast. An approximate resolution of 15 line pairs/mm (as derived from the resolution target image) was achieved without optimizing factors such as the optical properties and quality of the crystal, the depth of focus in the bulk of the medium, the carrier frequency of the grating, the relative intensities and wavelengths of the various beams, or the Bragg diffraction condition (discussed below). This spatial bandwidth is comparable to that obtained with a PROM [13] or a liquid crystal light valve [14].

A number of distinct factors influence the ultimate resolution achievable with the PICOC spatial light modulator. These factors can be classified as geometric, systemic, and materials-related in nature. Each of these resolution limitations are described in detail below.



(a)



(b)

Fig. 4 Photorefractive incoherent-to-coherent optical conversion of two binary transparencies: (a) spoke target, and (b) USAF resolution target.



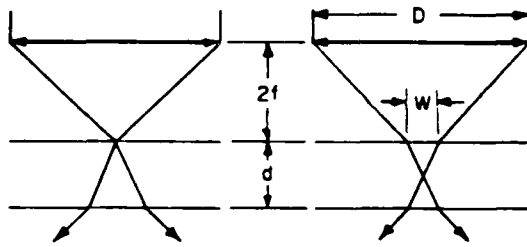
Fig. 5 Photorefractive incoherent-to-coherent optical conversion of a grey-level transparency.

Geometric resolution limitations derive principally from the incorporation of an incoherent imaging system in the four-wave mixing geometry, and from the finite crystal thickness required to create a volume holographic grating. These effects are illustrated in Figs. 6 and 7. Figure 6 describes the case for which $\alpha_s d \ll 1$, such that the induced holographic grating has finite amplitude throughout the volume of the crystal. As can be seen from the Figure, the optimum focal point occurs in the volume of the crystal (in the center when $\alpha_s d \ll 1$), and is not localized on the front surface of the crystal. The resolution will then be proportional to $(W/2)^{-1}$, which is in turn equal to $4n F\# / d$. In this expression, n is the refractive index of the electrooptic crystal, $F\#$ is the F-number of the input incoherent imaging system (assumed 1:1), and d is the crystal thickness. For example, for $n = 2.5$, $d = 1\text{mm}$, and an F-number of 5, the resolution limit is approximately 50 line pairs/mm. Figure 7 describes the case for which $\alpha_s d \gg 1$, such that the induced holographic grating has significant amplitude only within a thin layer of thickness $d_{\text{eff}} = \alpha_s^{-1}$. In this case, the resolution will be given by $4n F\# \alpha_s$. For $n = 2.5$, $d = 1\text{mm}$, $\alpha_s = 100\text{cm}^{-1}$, and an F-number of 5, the resolution limit is approximately 500 line pairs/mm.

Several systemic resolution constraints derive from the utilization of a volume hologram in the image conversion process. One such constraint is that the coherent grating spatial frequency must be significantly larger than the signal (incoherent image) bandwidth

in order to effect separation of the 0th and 1st order diffracted beam components. In addition, the range of spatial frequencies distributed about the coherent grating spatial frequency should be in an optimum region of the modulation transfer characteristic of the device.

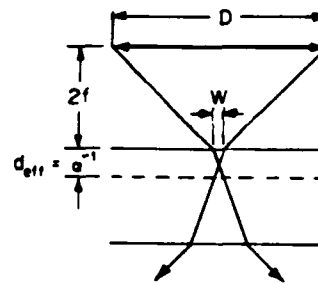
PHOTOREFRACTIVE INCOHERENT-TO-COHERENT
OPTICAL CONVERTER:
GEOMETRIC CONSTRAINTS



CASE I: $ad < 1$ $R \sim \frac{1}{(W/2)} = \frac{4nF\#}{d}$

EXAMPLE: FOR $n=2.5$, $d=1\text{mm}$, $F\#=5$,
 $R = 50$ line pairs/mm

PHOTOREFRACTIVE INCOHERENT-TO-COHERENT
OPTICAL CONVERTER:
GEOMETRIC CONSTRAINTS



CASE II: $ad \gg 1$ $R \sim \frac{1}{(W/2)} = 4nF\#a$

EXAMPLE: FOR $n=2.5$, $d=1\text{mm}$, $F\#=5$, $a=100\text{cm}^{-1}$
 $R = 500$ line pairs/mm

Fig. 6 Geometric constraints imposed on PICOC resolution by the finite F-number of the input optics, for the case $ad < 1$.

Fig. 7 Geometric constraints imposed on PICOC resolution by the finite F-number of the input optics, for the case $ad \gg 1$.

A third fundamental systemic constraint on the PICOC device resolution derives from Bragg detuning effects associated directly with the use of a volume hologram. This effect is strikingly illustrated in Figure 8, which is comprised of converted images of a 5 line pair/mm Ronchi ruling and their related coherent Fourier transforms, for two orthogonal orientations of the incoherent grating wavevector relative to the coherent grating wavevector. As can be seen from the Figure, a significant difference in resolution exists between cases in which the wavevector of the ruling (incoherent grating) is parallel or perpendicular to the coherently-written (holographic) grating. This difference derives principally from a wavevector matching condition, as shown in Figs. 9 and 10. In Fig. 9, the incoherent grating wavevector is parallel to the coherent grating wavevector, a condition achieved by symmetrically disposing the incident coherent beams about the normal to the crystal, while simultaneously arranging the incoherent imaging system such that its optical axis is parallel to the crystal normal. In this case, significant Bragg detuning occurs for even small incoherent grating wavevectors (resulting in angular deviations of the diffracted beam). In Fig. 10, the incoherent grating wavevector is arranged to lie tangent to the circle defined by the incident coherent grating wavevectors, such that a significantly increased angular deviation of the diffracted beam is allowed before significant Bragg detuning effects occur. Such a wavevector tangency condition is automatically satisfied when the incoherent grating wavevector is normal to the coherent grating wavevector (as it is in the orientation normal to the plane of incidence). For the arrangement described in Fig. 10, wavevector tangency is assured for both orientations, so that the horizontal and vertical resolutions become degenerate for this case. An equivalent condition is described in Ref. [7] for the case of distinct reading and writing beam wavelengths.

A fourth related systemic constraint involves the tradeoff between saturation diffraction efficiency and achievable resolution due to grating thickness. Within limits, increases in grating thickness increase the diffraction efficiency, but at the same time increase the Bragg detuning per unit angle. This condition creates a familiar resolution/sensitivity tradeoff situation in which one can be optimized at the expense of the other.

Additional resolution limitations derive from material parameter constraints that influence the physics of grating formation. This limitation is essentially described by the behavior of the diffraction efficiency as a function of the spatial frequency with the applied transverse electric field as a parameter [11]. The amplitude modulation transfer function of the grating itself depends on the spatial frequency dependence of the maximum space-charge field, which in turn depends on carrier mobilities and lifetimes, the applied field, the nature and distribution of traps in the crystal, and the trap occupancy (Fermi) level.

Optimization of the PICOC device resolution necessarily involves simultaneous satisfaction of all three types of constraints, and must be undertaken also in compromise with factors that directly affect the device sensitivity. In succeeding paragraphs, we

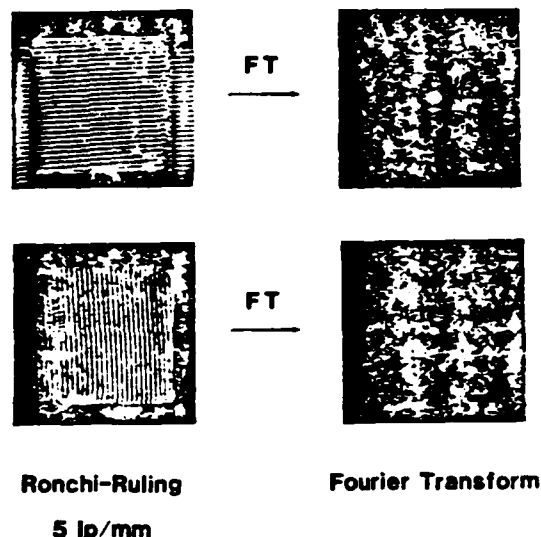


Fig. 8 Coherent replica images and corresponding Fourier transforms of a 5 line pair per mm Ronchi ruling after photorefractive incoherent-to-coherent conversion.

present a model that allows the device sensitivity to be understood and optimized. It should be noted here that amplification via energy transfer has been observed in the four-wave mixing configuration in BaTiO_3 [15], and in the two-wave mixing configuration in a BSO crystal [16, 17]. By utilizing this effect in the PICOC geometry described in this paper, incoherent-to-coherent conversion could be performed with improved energy sensitivity without sacrifice of resolution.

The following simple model describes the encoding of the incoherent information onto the coherent diffracted beam, by taking into account only the effective modulation ratio of the exposure in the bulk of the photorefractive material. The different intensities incident on the crystal are as shown in Fig. 2. The intensity pattern resulting from simultaneous illumination with the coherent writing beams and the incoherent image-bearing beam is given by Eqn. (1), which describes the case of a one-dimensional incoherent sinusoidal grating with a fundamental component at wavevector K_s . The space-charge field E_{sc} is related to the current density $J(x, t)$ and the free-carrier concentration $n(x, t)$ through the following system of equations [18]:

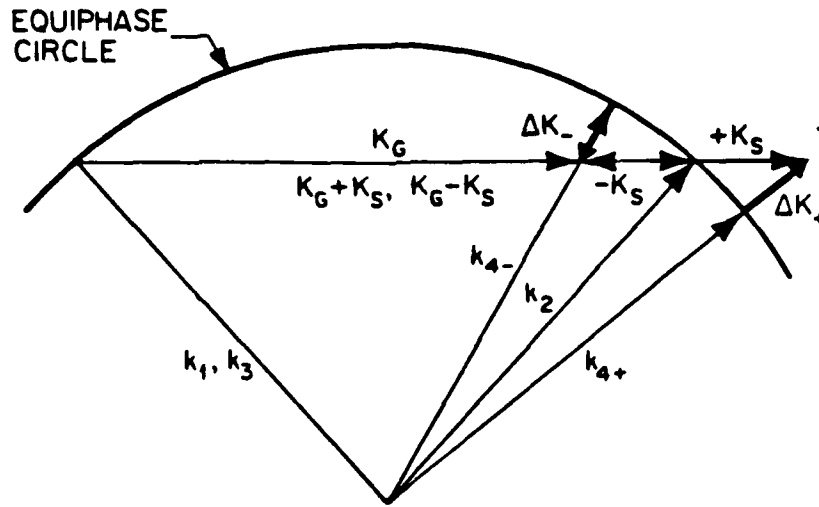


Fig. 9 Wavevector diagram of the photorefractive incoherent-to-coherent conversion process, for the case of collinear signal (K_S) and grating (K_G) wavevectors. Readout with a beam incident at wavevector k_3 results in diffracted beams at k_{4+} and k_{4-} , with corresponding Bragg mismatches ΔK_+ and ΔK_- .

Current density equation

$$J(x,t) = eD \frac{\partial n(x,t)}{\partial x} + e\mu n(x,t) [E_{sc}(x,t) + E_0] \quad (3)$$

Continuity equation

$$\frac{\partial n(x,t)}{\partial t} = g(x) - \frac{n(x,t) - n_d}{\tau} + \frac{1}{e} \frac{\partial J(x,t)}{\partial x} \quad (4)$$

Maxwell's equation (integral form)

$$E_{sc}(x,t) = -\frac{1}{\epsilon} \int_0^t J(x,t) dt + G(t) \quad (5)$$

$$\int_0^L E_{sc}(x,t) dx = 0 \quad (6)$$

in which $g(x)$ is the generation rate, n_d is the free-carrier concentration in the dark, E_0 is the applied transverse field, and $G(t)$ is a boundary condition-dependent constant of integration. If one assumes that the quantum efficiency ξ , which is the probability of creating a free electron by an absorbed photon of energy $\hbar\omega$, is not wavelength-dependent in the range of interest, the intensity distribution $I_T(x)$ gives rise to the free-carrier generation rate

$$g(x) = \frac{\alpha_G \xi}{\hbar\omega_G} I_G (1 + m_G \cos K_G x) \exp(-\alpha_G d) + \frac{\alpha_S \xi}{\hbar\omega_S} I_S (1 + m_S \cos K_S x) \exp(-\alpha_S d) \quad (7)$$

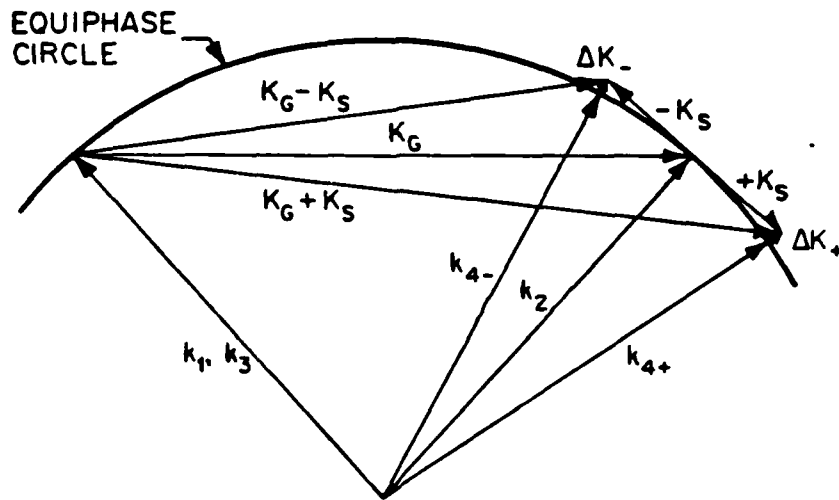


Fig. 10 Wavevector diagram of the photorefractive incoherent-to-coherent optical conversion process, for the case wherein the signal wavevector is tangent to the equiphase circle. Readout with a beam incident at wavevector k_3 results in diffracted beams at k_{4+} and k_{4-} , with significantly reduced Bragg mismatches ΔK_+ and ΔK_- .

in which d is the crystal thickness, and the intensity absorption coefficients are α_s at λ_s and α_G at λ_G . Since free-carrier redistribution will occur both by drift and diffusion, the resultant space-charge field will depend on both the carrier concentration $n(x)$ and its spatial gradient $\partial n/\partial x$. Solution of this system of equations is extremely complicated in general, and must be accomplished in the context of boundary conditions applicable to a given experimental situation. In the development that follows, several simplifying assumptions are made that are appropriate for the SEWM configuration in the steady state, and in the large transverse applied field limit (to enhance the saturation diffraction efficiency).

In the saturation regime (steady state), the current density is constant and the carrier concentration is given by

$$n(x) = n_d + \tau g(x) \quad (8)$$

in which the free-carrier lifetime τ is assumed to be the same for the writing and the erasure beams, and unaffected by the presence or absence of illumination. In the strong drift regime for which the applied field is much larger than the diffusion field ($E_0 \gg E_D$; $E_D = K_B T/e$), solution of the system of Eqns. (3-6) subject to the additional constraints expressed in Eqns. (7) and (8) yields the following expression for the space-charge field

$$E_{sc} = \frac{m_G^{\text{eff}} E_{DG} \sin K_G x + m_S^{\text{eff}} E_{DS} \sin K_S x}{1 + m_G^{\text{eff}} \cos K_G x + m_S^{\text{eff}} \cos K_S x} \quad (9)$$

$$+ E_0 \left[\frac{(1 - (m_G^{\text{eff}})^2)^{1/2}}{1 + m_G^{\text{eff}} \cos K_G x + m_S^{\text{eff}} \cos K_S x} - 1 \right]$$

with

$$m_G^{\text{eff}} = \frac{m_G}{1 + \frac{\alpha_S \lambda_S}{\alpha_G \lambda_G} \frac{I_S}{I_G} \exp [(\alpha_G - \alpha_S) d]} \quad (10)$$

and

$$m_S^{\text{eff}} = \frac{m_S}{1 + \frac{\alpha_S \lambda_S}{\alpha_G \lambda_G} \frac{I_S}{I_G} \exp [(\alpha_G - \alpha_S) d]} \quad (11)$$

Assuming further that the effective fringe modulation ratios (m_G^{eff} , m_S^{eff}) are small and that $E_{DS} \ll E_{DG} \ll E_0$, the space-charge field amplitudes at each harmonic component of interest are given by:

$$E_{sc}(K_G) = -m_G^{\text{eff}} E_0 \quad (12)$$

$$E_{sc}(K_G + K_S) = m_G^{\text{eff}} m_S^{\text{eff}} E_0 \quad (13)$$

$$E_{sc}(K_G - K_S) = m_G^{\text{eff}} m_S^{\text{eff}} E_0 \quad (14)$$

The refractive index modulation induced along a principal axis by the electrooptic effect is then given by

$$\begin{aligned} \Delta n(x) &= \frac{1}{2} n_0^3 r_{41} E_{sc}(x) \\ &= -\frac{1}{2} n_0^3 r_{41} m_G^{\text{eff}} E_0 [\cos K_G x - m_S^{\text{eff}} \cos(K_G + K_S)x - m_S^{\text{eff}} \cos(K_G - K_S)x] \end{aligned} \quad (15)$$

or, by combining terms,

$$\Delta n(x) = -\frac{1}{2} n_0^3 r_{41} m_G^{\text{eff}} E_0 [1 - 2m_S^{\text{eff}} \cos K_S x] \cos K_G x. \quad (16)$$

For small diffraction efficiencies and low spatial frequencies K_S , the diffracted intensity pattern in the first order of the carrier grating can be expressed by the relation

$$I_4(x) = I_3 \left(\frac{\pi d n_0^3 r_{41} m_G^{\text{eff}} E_0}{\lambda_G \cos \theta} \right)^2 \left(1 - 4m_S^{\text{eff}} \cos K_S x \right) \quad (17)$$

in the Kogelnik formulation [19]. This expression demonstrates that the information contained in the incoherent incident intensity distribution $I_S(x)$ (see Eqn. (1)) is transferred onto the coherent diffracted beam, performing an incoherent-to-coherent conversion in which the original contrast is reversed.

For the case of uniform erasure ($K_S = 0$), the diffracted beams I_{11} and I_{10} are degenerate (see Fig. 3). For the case of modulated erasure ($K_S \neq 0$), this degeneracy is broken and separate beams are observed. The diffraction efficiencies in these two cases are given by:

$$\eta(I_{10}) = [m_G^{\text{eff}}(R)]^2 E_0^2 \quad (18)$$

$$\eta(I_{11}) = [m_G^{\text{eff}}(R) m_S^{\text{eff}}(R)]^2 E_0^2 \quad (19)$$

where R is the ratio of the incoherent to the coherent beam intensities ($R = I_S/I_G$). Figures 11 and 12 illustrate the normalized diffraction efficiency of the I_{10} beam for three typical erasure wavelengths (λ_S). The parameter A_S is the ratio of the erasure beam to the writing beam intensity absorption coefficients ($A_S = \alpha_S/\alpha_G$). It is interesting to note that the maximum energy sensitivity does not generally occur for wavelength-matching ($\lambda_G = \lambda_S$). In fact, two limiting situations can be recognized. For a thin crystal (Fig. 11), the optimum sensitivity is obtained with a strongly absorbed erasure beam because the generation rate increases with the ratio A_S , while the attenuation factor [$\exp(\alpha_G - \alpha_S)d$] has almost no effect on the effective modulation m_G^{eff} . Conversely, for a thick crystal (Fig. 12), the exponential term dominates and the optimum sensitivity is achieved when the incoherent source has the same wavelength as the laser with which the carrier grating is recorded. This is intuitively appealing since the phase grating is most effectively erased throughout the volume of the crystal for the case of wavelength matching.

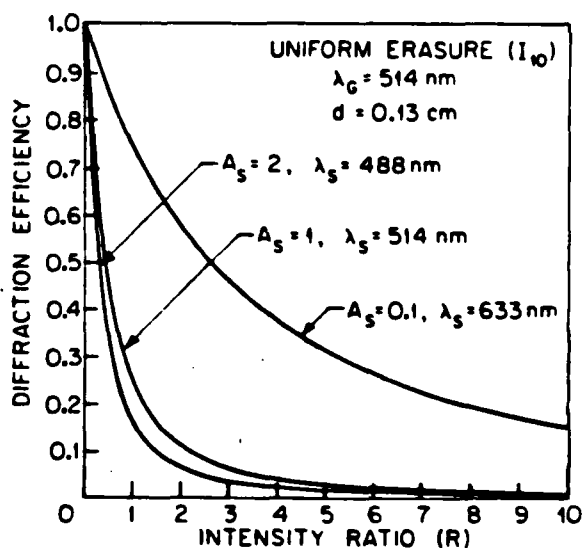


Fig. 11 Normalized diffraction efficiency of the I_{10} beam (Uniform Erasure) as a function of the intensity ratio (R) for a 0.13 cm thick crystal. Three signal beam (S) wavelengths (λ_S) with corresponding absorption coefficients (A_S) are used to erase the coherent grating written at wavelength λ_G .

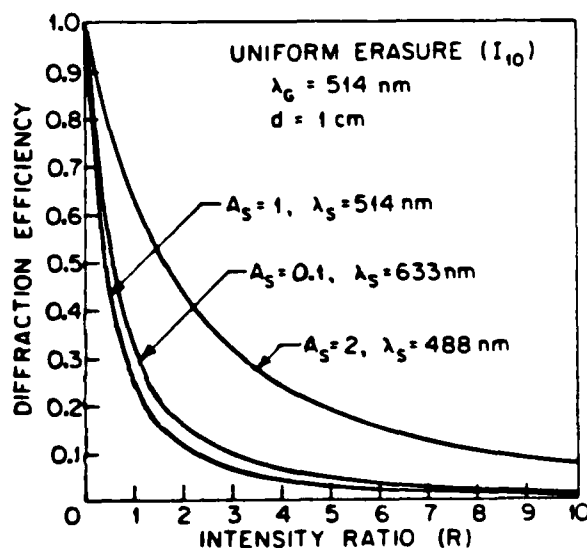


Fig. 12 Normalized diffraction efficiency of the I_{10} beam (Uniform Erasure) as a function of the intensity ratio (R) for a 1.0 cm thick crystal. Three signal beam (S) wavelengths (λ_S) with corresponding absorption coefficients (A_S) are used to erase the coherent grating written at wavelength λ_G .

The normalized diffraction efficiency for the beam I_{11} in the case of modulated erasure is shown in Fig. 13. As predicted by Eqn. (19), a maximum exists for $R_{eff} = 1$ where:

$$R_{eff} = R \frac{a_S \lambda_S}{a_G \lambda_G} \exp(a_G - \lambda_S) d \quad (20)$$

The existence of such a maximum is explained by the following. It is clear that when $R_{eff} = 0$, there is no diffraction in the I_{11} direction. In addition, when R_{eff} is large, the uniform component of the incoherent exposure almost completely erases the carrier grating and the diffracted intensity vanishes. Furthermore, a small intensity ratio ($R_{eff} \sim 1$) generates the maximum signal (I_{11}), but the uniform erasure (I_{10}) is weak and as a result the overall contrast is low. Conversely, a large ratio ($R_{eff} \gg 1$), decreases the modulation m_{eff} , but at the same time diffraction in the beam I_{10} is minimized so that the overall contrast is enhanced at the expense of sensitivity. In Figs. 4 and 5, R was set equal to 10

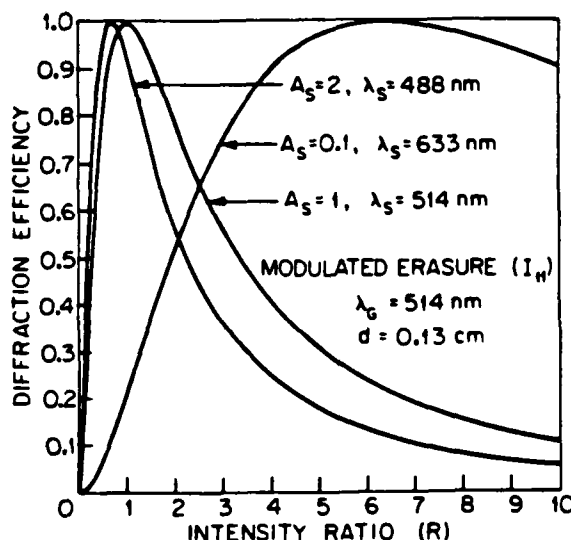


Fig. 13 Normalized diffraction efficiency of the I_{11} beam (Modulated Erasure) as a function of the intensity ratio (R) for a 0.13 cm-thick crystal. Three signal beam (S) wavelengths (λ_S) with corresponding absorption coefficients (A_S) are used to erase the coherent grating written at wavelength λ_G .

in order to maximize the contrast in the negative imaging mode. However, a number of specialized readout techniques (e.g. Schlieren, phase contrast) could be employed to simultaneously optimize the signal intensity and the contrast ratio.

To verify the essential predictions of this simple model described above, the experimental arrangement shown in Fig. 14 has been utilized. The carrier grating is recorded in the BSO crystal at $\lambda_G = 514 \text{ nm}$ and read out with a He-Ne laser at $\lambda_R = 633 \text{ nm}$ incident at the Bragg angle. The incoherent grating is formed with a Michelson interferometer (M_4, M_5) using the blue line of the argon laser ($\lambda_S = 488 \text{ nm}$). The far-field diffraction pattern is located in the Fourier plane of the lens L . The results of theoretical calculations and experimental measurements of the normalized diffraction

The general temporal response behavior of the I_{11} (information-bearing) diffracted beam is shown in Fig. 17 for the two cases $R = 1.5$ and $R = 5.0$. In this experiment, the coherent grating is established and in the saturation regime at time $t = 0$, at which point the incoherent erasure beam is allowed to expose the crystal. For a small ratio R ($I_S = 0.6$ mW/cm² in Fig. 17), the diffraction efficiency increases steadily with time until it reaches a maximum. However, for a strong incoherent beam ($I_S = 2$ mW/cm² in Fig. 17), a transient effect appears. Initially, the beam I_{10} is quite intense and diffracts from the composite grating at $K_G + K_S$ to generate a rapid rise in the amplitude of I_{11} , but the uniform part of the incoherent illumination simultaneously erases the coherent carrier grating and hence the diffraction efficiency decreases to a small steady-state value. The time constant for this particular set of experimental parameters is in the range 0.5-1.5 sec., although the device response time can be decreased by increasing the intensity of the carrier grating writing beams. In addition, the response time is strongly dependent on a number of intrinsic and extrinsic materials parameters. The full range of variation possible for the temporal behavior characteristic is not firmly established at this point.

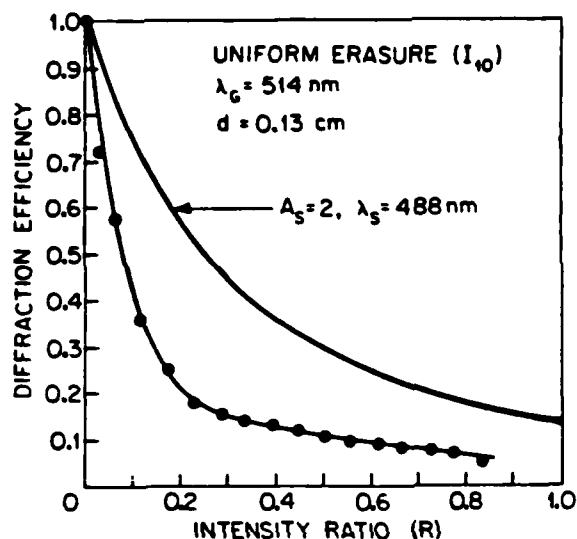


Fig. 15 Normalized experimental diffraction efficiency of the I_{10} beam (Uniform Erasure) as a function of the intensity ratio (R). The corresponding theoretical prediction is also shown for comparison.

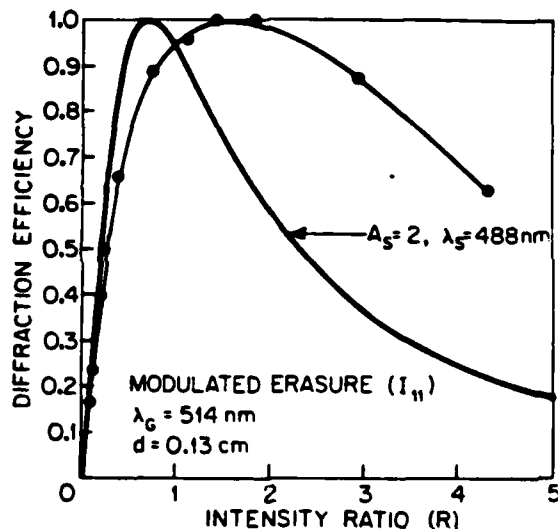


Fig. 16 Normalized experimental diffraction efficiency of the I_{11} beam (Modulated Erasure) as a function of the intensity ratio (R). The corresponding theoretical prediction is also shown for comparison.

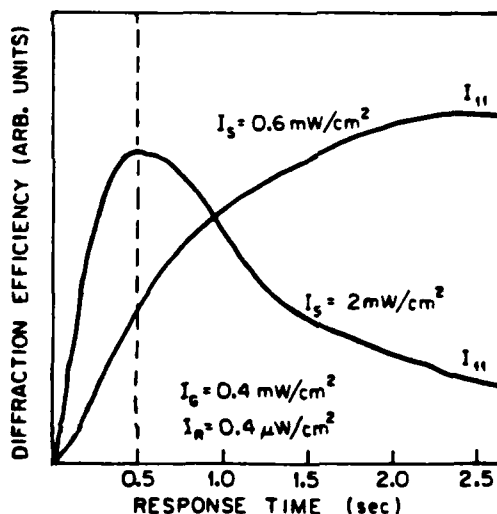


Fig. 17 Measured diffraction efficiency of the I_{11} beam (Modulated Erasure) as a function of response time for two values of the signal intensity.

In conclusion, we have demonstrated the feasibility of real-time incoherent-to-coherent conversion utilizing phase conjugation in photorefractive BSO crystals. The physical basis of this effect is explained with a simple model which takes into account the influence of the effective modulation ratio on the diffraction efficiency of the different diffracted beams. Although the results are preliminary, the device which is proposed in this paper has potential for incoherent-to-coherent conversion with high resolution, which can be realized by optimizing the wavevector matching condition and the relative intensities and wavelengths of the various beams. In addition, the PICOC device is quite attractive from considerations of low cost, ease of fabrication, and broad availability. With such a device, numerous optical processing functions can be directly implemented that utilize the flexibility afforded by the simultaneous availability of incoherent-to-coherent conversion and volume holographic storage.

ACKNOWLEDGEMENTS

The authors would like to thank F. Lum, D. Seery, and M. Garrett for their technical assistance. This research was supported in part at USC by the Air Force Systems Command (RADC) under Contract No. F19628-83-C-0031, the Defense Advanced Research Projects Agency, the Joint Services Electronics Program, and the Army Research Office; and at Caltech by the Air Force Office of Scientific Research and the Army Research Office.

REFERENCES

- [1] J. O. White and A. Yariv, Appl. Phys. Lett. 37, 5 (1980).
- [2] A. Marrakchi, J. P. Huignard and J. P. Herriau, Opt. Comm. 34, 15 (1980).
- [3] G. J. Dunning and R. C. Lind, Opt. Lett. 7, 558 (1982).
- [4] A. R. Tanguay, Jr., Proc. ARO Workshop on Future Directions for Optical Information Processing, Texas Tech University, Lubbock, Texas, May, 1980, 52-77 (1981).
- [5] D. Casasent, Proc. IEEE, 65, 143 (1977).
- [6] Y. Shi, D. Psaltis, A. Marrakchi and A. R. Tanguay, Jr., Appl. Opt. 22, 3665 (1983).
- [7] A. A. Kamshilin and M. P. Petrov, Sov. Tech. Phys. Lett. 6, 144 (1980).
- [8] D. L. Staebler and J. J. Amodei, J. Appl. Phys. 43, 1042 (1972).
- [9] M. Peltier and F. Micheron, J. Appl. Phys. 48, 3683 (1977).
- [10] R. Grousson and S. Mallick, Appl. Opt. 19, 1762 (1980).
- [11] J. P. Huignard, J. P. Herriau, G. Rivet and P. Gunter, Opt. Lett. 5, 102 (1980).
- [12] J. P. Herriau, J. P. Huignard and P. Aubourg, Appl. Opt. 17, 1851 (1978).
- [13] Y. Owechko and A. R. Tanguay, Jr., Opt. Lett. 7, 587 (1982).
- [14] P. Aubourg, J. P. Huignard, M. Hareng and R. A. Mullen, Appl. Opt. 21, 3706 (1982).
- [15] J. Feinberg and R. W. Hellwarth, Opt. Lett. 5, 519 (1980).
- [16] J. P. Huignard and A. Marrakchi, Opt. Comm. 38, 249 (1981).
- [17] H. Rajbenbach, J. P. Huignard and B. Loiseaux, Opt. Comm. 48, 247 (1983).
- [18] M. G. Moharam, T. K. Gaylord and R. Magnusson, J. Appl. Phys. 50, 5642 (1979).
- [19] H. Kogelnik, Bell Syst. Tech. J. 18, 2909 (1969).

Bias-free time-integrating optical correlator using a photorefractive crystal

Demetri Psaltis, Jeffrey Yu, and John Hong

An acoustooptic time-integrating correlator is demonstrated using a photorefractive crystal as the time-integrating detector.

I. Introduction

Time integration¹ has proved to be a powerful technique in optical signal processing and has been used in a wide variety of architectures. A major drawback of time-integrating processors is the buildup of bias in addition to the signal. This occurs because the photo-generated charge that is integrated on the detector is proportional to the intensity of the optical signal which makes it necessary to represent bipolar signals on a bias. The effective system dynamic range at the output is given by $DR' = DR [SBR/(1 + SBR)]$ where DR is the dynamic range of the output detector and SBR is the signal-to-bias ratio on the detector.² In most cases of interest, the SBR is much smaller than unity and thus the added bias significantly reduces the usable dynamic range of the system.

The most frequently used method for separating the signal from the bias involves placing the signal on a spatial carrier and then electronically filtering the output of the integrator. This method of bias removal, however, does not solve the dynamic range problem since the processing is done after the detection of the signal. Also, an additional constraint is placed on the resolution of the detector, since the pixel separation must be less than one-half of the period of the carrier being recorded, which will result in a significant reduction in the available space-bandwidth product at the output.

In this paper a new method for performing time-integrating correlation is described using a photorefractive bismuth silicon oxide (BSO) crystal as the time-integrating element. The correlation is formed

on a spatial carrier in the crystal and read out with an auxiliary beam. Since only the signal recorded on a spatial carrier is stored in the photorefractive crystal, the diffracted light that is detected contains the correlation information without the bias. The bias does not reduce the dynamic range of the output detector used for final readout, but rather the diffraction efficiency of the BSO crystal. In addition, the resolution of the BSO crystal is very much higher than that of a CCD detector, allowing the correlation of very high space-bandwidth signals to be formed on a carrier. Finally, since the result of the time-integrating correlator is read out optically, the output can be easily interfaced with other optical systems, thus making new architectural designs possible.

In Sec. II, the theory of optical recording in photorefractive crystal is reviewed and extended to the use of photorefractive crystals as time-integrating elements. The architecture and experimental results are described in Sec. III. Dynamic range, linearity, system limitations, and other performance aspects are discussed in Sec. IV.

II. Photorefractive Crystals as Time-Integrating Optical Detectors

When a photorefractive BSO crystal is illuminated by an intensity grating, electrons are excited from traps into the conduction band. These charges migrate due to diffusion and drift from an externally applied electric field and then recombine in dark regions, creating a spatially varying internal space-charge field. This field modifies the index of refraction in the crystal through the linear electrooptic effect and, as a result, a holographic phase grating is recorded in the crystal. Grating formation in photorefractive media has been extensively studied and modeled.^{3,4} We will show here that the photorefractive crystal acts as a time-integrating element.

Let the intensity incident on the crystal be as follows:

The authors are with California Institute of Technology, Department of Electrical Engineering, Pasadena, California 91125.

Received 20 June 1985.

0003-6935/85/223860-06\$02.00/0.

© 1985 Optical Society of America.

$$I(x,t) = \begin{cases} I_0 + \text{Re}[I_1(x,t) \exp(ikx)] & \text{for } t > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Assuming that self-diffraction effects are negligible and that the spatial variations of $I_1(x,t)$ are small compared to the grating frequency k , the intensity of the light that is diffracted when the crystal is illuminated by a readout beam can be shown to be⁵

$$I_{\text{out}}(x,t) = \left| \frac{K_1}{\tau} \exp(-t/\tau) \int_0^t \frac{I_1(x,t')}{I_0} \exp(t'/\tau) dt' \right|^2 I_R \quad (2)$$

I_R is the intensity of the readout beam and K_1 is a complex constant involving the material parameters of the crystal, the grating frequency, and the applied electric field. τ is the complex time constant of the space-charge field and is given by³ $\tau = K_2/I_0$. K_2 is also a complex constant that depends on the photorefractive material used and the experimental conditions. I_0 is the average light intensity incident on the crystal during exposure.

If $I_1(x,t)$ is expanded into its temporal Fourier components,

$$I_1(x,t) = \int_{-\infty}^{\infty} I_1(x,\omega) \exp(i\omega t) d\omega,$$

then the output intensity for $t \gg \tau$ can be written as follows:

$$I_{\text{out}}(x,t) \approx \left| \frac{K_1}{\tau I_0} \int_{-\infty}^{\infty} \frac{I_1(x,\omega)}{(1/\tau + i\omega)} \exp(i\omega t) d\omega \right|^2 I_R \quad (3)$$

The above is recognized to be a low pass filter with cutoff frequency $1/|\tau|$, which is approximately equivalent to the output of a sliding window integrator, with integration time τ . Thus,

$$I_{\text{out}}(x,t) \approx \left| \frac{K_1}{\tau} \int_0^t \frac{I_1(x,t')}{I_0} dt' \right|^2 I_R \quad (4)$$

Hence, the output intensity can be treated as the square of the normalized integration of the signal I_1 . An interesting observation is that, if $I_1(x,t)$ were independent of time, the output intensity depends only on the ratio of the modulated intensity I_1 to the dc intensity I_0 .

III. Experimental Demonstration

A schematic diagram of the experimental system is shown in Fig. 1. The input electrical signals are mixed with the center frequency of the acoustooptic devices (AODs) and fed into the AODs. The first AOD is illuminated by a collimated wave and the upshifted diffracted order is imaged onto the second AOD, then reimaged onto the photorefractive BSO crystal. The second AOD is oriented such that the acoustic signal is counterpropagating with respect to the image of the acoustic signal from the first AOD. The undiffracted light transmitted through the first AOD is incident at the Bragg angle of the second AOD. The upshifted diffracted order of the second AOD is also imaged onto the BSO crystal. The undiffracted light is spatially filtered before reaching the BSO crystal. In this arrangement, the AODs are parallel to each other, but the diffracted orders propagate at an angle with respect to each other even though both diffracted beams are temporally upshifted. This causes the signals from the two AODs to interferometrically record the correlation signal on the BSO crystal at a high spatial frequency. Let the inputs to the AODs be $v_1(t) = a(t) \exp(i\omega_0 t)$ and $v_2(t) = b(t) \exp(i\omega_0 t)$, where $\omega_0/2\pi$ is the center frequency of the AOD. The intensity incident on the photorefractive crystal is

$$\begin{aligned} I(x,t) &= |a(t - x/v) \exp(i\gamma x) + b(t + x/v) \exp(-i\gamma x)|^2 \\ &= |a(t - x/v)|^2 + |b(t + x/v)|^2 \\ &\quad + 2 \text{Re}[a(t - x/v)b^*(t + x/v) \exp(i2\gamma x)], \end{aligned} \quad (5)$$

where v is the acoustic velocity of the AOD and $\gamma = \omega_0/v$. We will treat the case where $|a(t)|^2$ and $|b(t)|^2$ can both be approximated as constants, as is the case for FM signals. Then, the intensity pattern can be separated into a dc term I_0 and a signal term $I_1(x,t)$ modulating a spatial carrier $\cos(2\gamma x)$ in the form of Eq. (1). This intensity pattern results in the formation of a hologram on the photorefractive crystal as described in the previous section. The hologram is read out with an auxiliary beam and is imaged onto a charge coupled device (CCD) detector for readout.

If the assumption is valid that $I_1(x,t)$ has spatial frequencies which are small compared with the carrier frequency 2γ , we can use Eq. (4) to obtain an expression for the output intensity detected by the CCD:

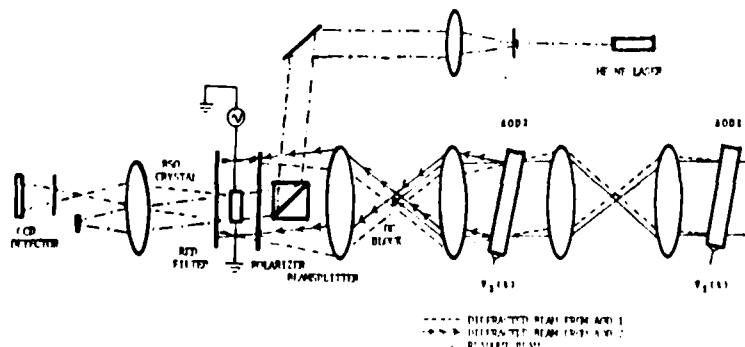


Fig. 1. Optical setup of the photorefractive bias removal correlator.

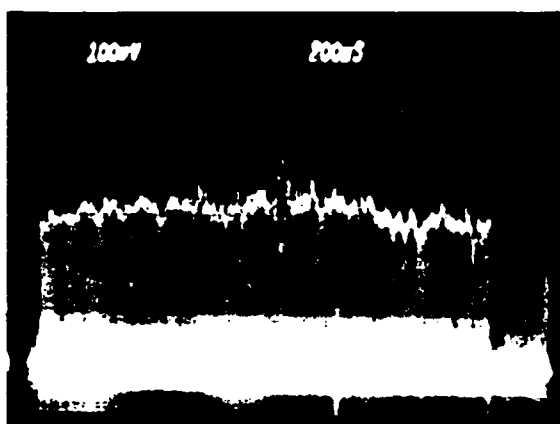


Fig. 2. Output of a standard time-integrating correlator without noise.



Fig. 3. Output of the bias removal correlator without noise.

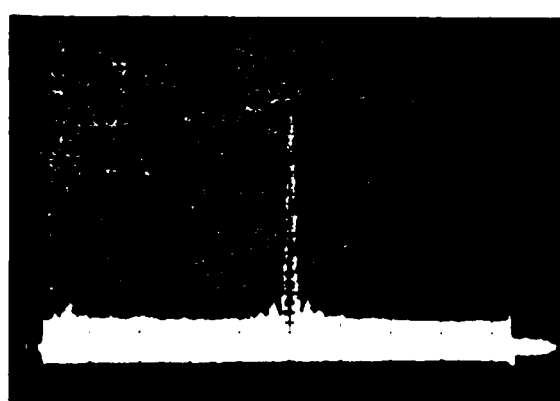


Fig. 4. Output of the bias removal correlator with a signal-to-noise ratio of 0 dB.

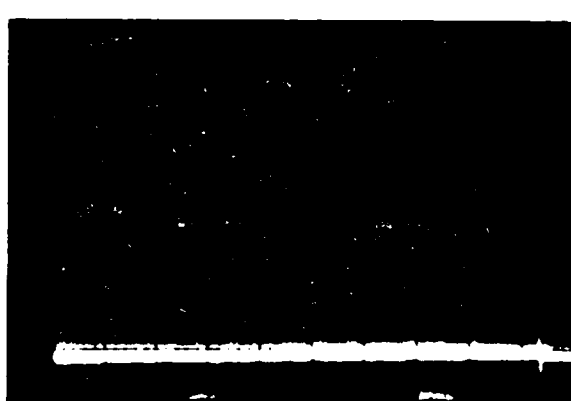


Fig. 5. Output of the bias removal correlator with a signal-to-noise ratio of -10 dB.

$$I_{\text{out}}(x) \approx I_R \left| \frac{K_1}{\tau} \right|^2 \left| \int_t^{t+\tau} \frac{a(t-x/v)b^*(t+x/v)}{|a|^2 + |b|^2} dt \right|^2,$$

and by defining variable $t_1 = t - x/v$,

$$I_{\text{out}}(x) \propto \left| \int_{t-x/v}^{t+x/v} a(t_1)b^*(t_1 + 2x/v) dt_1 \right|^2. \quad (6)$$

Hence the system produces the magnitude square of the correlation between the signals $a(t)$ and $b(t)$ integrated over a finite interval τ .

Flint glass acoustooptic cells driven at a center frequency of 70 MHz were used in the experiment. A symmetric linear chirp signal with bandwidth $\Delta f = 5$ Mhz was fed into each cell to produce the autocorrelation peak. The Bragg angle of the AODs was 0.2° , which corresponded to a grating frequency equal to 35 lines/mm in the BSO crystal.

The BSO crystal used in the experiment was cut in the $\langle 110 \rangle$ direction and measured $15 \times 15 \times 2$ mm. An external electric field of 7 kV/cm was applied in the $\langle 001 \rangle$ direction of the crystal which was also the direction of the grating vector.

The correlation was recorded on the crystal with an argon laser at a wavelength of 514 nm with average intensity equal to $1 \mu\text{W}/\text{cm}^2$. The correlation was read out with a He-Ne laser ($\lambda = 633$ nm) with $150\text{-}\mu\text{W}/\text{cm}^2$ intensity. Cylindrical lenses (not shown in Fig. 1) were used to expand the output of the AODs thereby illuminating the full aperture of the BSO crystal and also to focus the diffracted light onto a 1-D CCD.

The output signal-to-bias ratio of a conventional time-integrating correlator is reduced when the levels of the two signals are unequal and/or if there is additive noise present in the system. Both conditions were simulated experimentally. Noise was simulated by adding a 70-MHz signal to the input of one of the AODs. The output of a standard time-integrating correlator (i.e., the correlation formed directly on the CCD) for the noise-free case and equal amplitude signals is shown in Fig. 2. This condition provides the maximum signal-to-bias ratio for the system. We can

see in Fig. 2 that there is still a strong bias term added to the correlation peak. The correlation produced by temporally integrating on the photorefractive crystal is shown in Fig. 3. In this case, all the bias due to temporal integration is removed, and any residual bias is due entirely to dark current from the CCD. The outputs of the bias removal correlator with input signal-to-noise ratios of 0 and -10 dB are shown in Figs. 4 and 5, respectively. Again, bias levels which appear in the figures were entirely due to the integration of dark current in the output detector. In practice, the detector dark current can be minimized by increasing the intensity of the readout beam, thereby decreasing the required integration time of the output CCD detector and/or cooling the detector.

III. Performance

The experimental results described in the previous section show a dramatic qualitative improvement in the correlation that is obtained when the photorefractive crystal is used instead of the CCD. In this section we examine certain characteristics of this method which are useful for quantitatively evaluating its performance. Specifically, we examine the linearity, integration time, dynamic range, and sensitivity of the correlator.

A. Linearity

In a conventional time-integrating correlator (coherent or incoherent), the output correlation is basically proportional to the signals applied to the AODs. Nonlinearities occur only when we exceed the linear dynamic range of the devices used, i.e., if the diffraction efficiency of the AOD exceeds several percent or the integrating detector is driven to saturation. In the photorefractive time-integrating processor, the output intensity is a nonlinear, monotonically increasing function of the input voltage. The nonlinearity arises because of the square-law detection at the final readout stage and the recording mechanism in the photorefractive crystal. The nonlinear relationship is now studied analytically and experimental verification of the theoretical results is presented.

Let $v_1(t) = s(t)$ be a fixed reference signal and $v_2(t) = as(t)$ be an input signal of varying amplitude ($0 < a < 1$). Since the correlation term contains spatial frequencies which are much lower than the grating frequency, near the correlation peak ($x = 0$) the intensity incident on the photorefractive crystal is

$$I(x,t) = (1 + a^2 + 2a \cos kx) |s(t)|^2.$$

Using Eq. (1), the output intensity at the CCD is proportional to

$$I_{out} \propto \left| \frac{2a}{1 + a^2} \right|^2.$$

The modulation depth of the intensity incident on the BSO crystal is

$$m = \frac{2a}{1 + a^2},$$

and hence

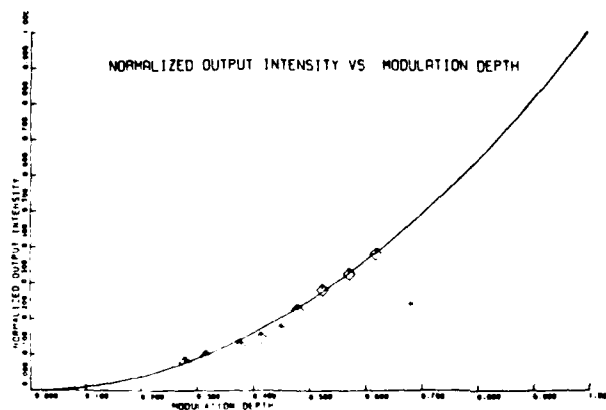


Fig. 6. Normalized output intensity vs modulation depth.

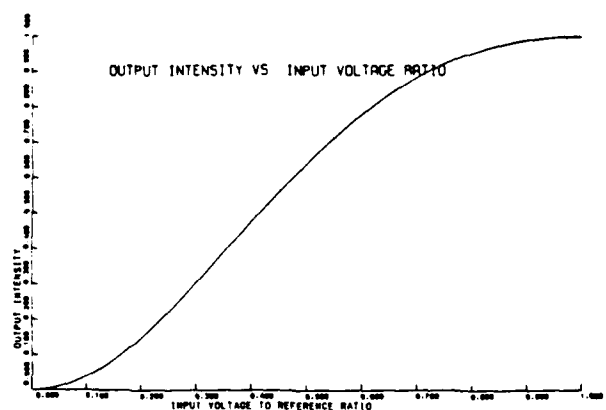


Fig. 7. Theoretical plot of output intensity vs input voltage ratio.

$$I_{out} \propto m^2 \approx 4a^2/(1 + a^2). \quad (7)$$

Figure 6 is a graph of the output intensity at the correlation peak vs the modulation depth incident on the crystal. The experimental result is in excellent agreement with the square-law relationship predicted by Eq. (7).

A plot of the output intensity as a function of the amplitude of the input signal a is shown in Fig. 7. The nonlinear relationship between the input and output signals is generally a disadvantage since the scaling of signals of varying amplitudes will be nonlinear. This, however, will not cause a problem if the correlator is used only as a signal detection device, since correlation peaks will still be discernible and only the threshold level need be adjusted accordingly to maximize the probability of detection.

B. Integration Time

In a conventional time-integrating correlator, the integration time is limited by the dark current buildup on the output detector, typically up to several hundred milliseconds. When the photorefractive crystal is

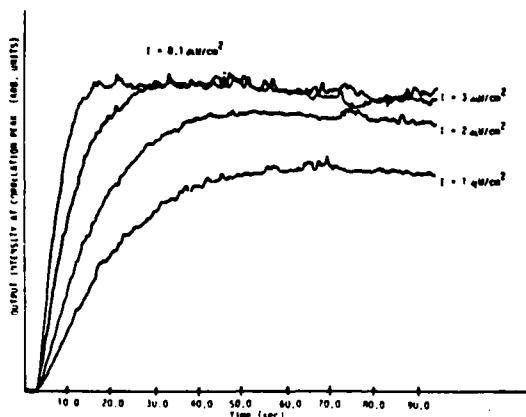


Fig. 8. Output intensity at correlation peak vs time as a function of different average incident intensities.

used, the integration time is determined by the rise time of the internal space-charge field which can easily be made much longer. The correlation can be read out at any rate that is convenient by an auxiliary detector array.

The integration time is approximately equal to $|\tau|$, where

$$|\tau| = \left| \frac{K_2}{I_0} \right|. \quad (8)$$

Hence, the integration time of the bias removal correlator can be controlled by varying the writing intensity. This control is important since the integration time can be matched to the length of the reference signal thereby increasing the probability of detection of a weak signal.

The time response of the correlation peak for different values of average incident intensity is shown in Fig. 8. Figure 9 is a plot of intensity vs the inverse of the experimentally observed rise time. There is excellent agreement between the experiment and Eq. (8).

The integration time, however, has a finite range over which it can be adjusted. The maximum integration time is limited by the thermal effects in the crystal. If the rate at which carriers are generated thermally becomes comparable with the rate at which they are photogenerated, the modulation depth of trap density will be reduced. As a result, the diffraction efficiency of the grating will decrease. In practice, the minimum integration time is limited by the maximum light intensity that is available for recording. The integration time can be reduced to 30 msec if the incident intensity is made equal to 18 mW/cm². This power level, however, is simply not practical for most applications.

C. Dynamic Range and Sensitivity

Since the output of the bias removal correlator is presented without bias, the output dynamic range of the system is essentially equal to the dynamic range of the readout detector array. To characterize the per-

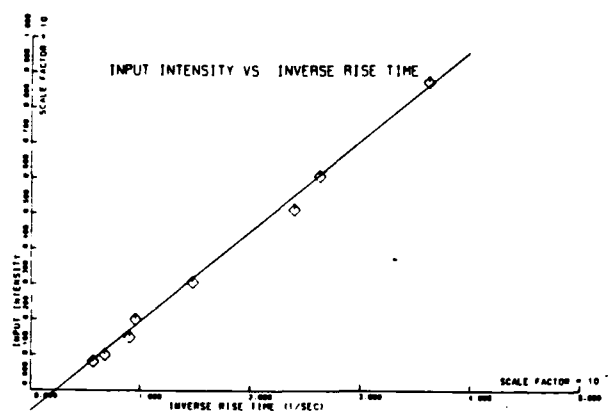


Fig. 9. Inverse of the rise time vs average incident intensity.

formance of the system we need to determine how the input signal levels are mapped to this output dynamic range. Let the dynamic range of the photorefractive crystal be defined as $DR_{BSO} = m_{\max}/m_{\min}$, where m_{\max} is the maximum modulation depth ($m_{\max} = 1$), and m_{\min} is the minimum modulation depth for which a diffracted signal is detectable above the output scatter and noise level of the system.

Given two input signals $v_1(t) = as(t)$ and $v_2(t) = s(t)$, the modulation depth of the light incident on the crystal is $m = 2a/(a^2 + 1)$. Thus, the minimum detectable input signal is given by $a_{\min} = m_{\min}/2 = 1/DR_{BSO}$. The useful range over which a can vary is limited by DR_{BSO} . From a_{\min} , one can define an input dynamic range given by $DR_{\text{input}} = 1/a_{\min}^2 = 4/m_{\min}^2$. The most important parameter in determining the system dynamic range is the minimum detectable modulation depth m_{\min} . Experimentally, we measured the dynamic range to be equal to 23 dB. This corresponds to a minimum modulation depth of 0.142. We expect that through careful design this can be substantially improved. However, all the mechanisms that determine m_{\min} are not fully understood. It is believed that besides detector noise and scattering from the crystal, the modulation depth is limited by thermal effects in the material and shot noise arising from the internal currents.

Another important aspect of the correlator system is its sensitivity or the minimum signal-to-noise ratio that is detectable. This parameter is also determined by the minimum detectable modulation depth, m_{\min} . Given a reference signal $v_1(t) = as(t)$ and an input signal contaminated by additive noise, $v_2(t) = bs(t) + n(t)$, the modulation depth of the intensity incident on the crystal is

$$m = \frac{2ab|s(t)|^2}{(a^2 + b^2)|s(t)|^2 + \sigma_n^2}.$$

The reference level which maximizes m is given by $a = (b^2 + \sigma_n^2/|s(t)|^2)^{1/2}$, corresponding to a modulation depth of

$$m = \frac{b}{(b^2 + \sigma_n^2 / |s(t)|^2)^{1/2}}$$

In practice, optimizing the reference level can easily be achieved by setting the power of the reference equal to the total average power of the input signal.

Normalizing the signal and noise terms such that $|s(t)|^2 = \sigma_n^2 = 1$ we obtain

$$m = \frac{b}{(b^2 + 1)^{1/2}}$$

Thus, the minimum input SNR that produces a detectable correlation peak at the output is $(S/N)_{\min} = (b^2 / \sigma_n^2)_{\min} \approx m_{\min}^2$.

From the experimentally measured value of m_{\min} , the correlator should have had a sensitivity of -17 dB. However, experimental results showed a sensitivity of -14 dB.

IV. Conclusion

The photorefractive time-integrating processor that has been described has several advantageous features: bias removal, increase in the output space-bandwidth product, and the ability to directly interface the result of the time-integrating processor with other optical systems. Bias-free correlation is desirable because it allows us to increase the dynamic range and hence the sensitivity of time-integrating processors. In the implementation described in this paper, however, the square-law detection at the output reduces the available overall dynamic range. A definite improvement in dynamic range can be obtained if the correlation that is formed in the photorefractive crystal is interferometrically detected on the output detector. Another limitation of the system described here is the long integration time (several seconds). In some applications this long integration time is desirable and could

result in extremely good sensitivity (detection of signals with very low SNR). However it is certainly desirable to be able to decrease the integration time to several milliseconds. This could be accomplished by increasing the optical power of the writing beams, but this is in general an impractical solution. Another limitation of this technique is the relatively low diffraction efficiency that is obtained with BSO crystals (2-3%), which reduces the overall light efficiency. Materials with higher electrooptic coefficients, such as barium titanate, can provide better efficiency; however, the time constant obtained with this particular material is much longer than that of BSO. New photorefractive materials currently being developed showing promise of a large improvement in optical sensitivity as well as higher electrooptic coefficients may provide a substantial improvement in performance and, specifically, reduce the total optical power that is required.

This work is supported by the Air Force Office of Scientific Research and the Army Research Office.

References

1. R. A. Sprague and C. L. Koliopoulos, "Time Integrating Acoustooptic Correlator," *Appl. Opt.* 15, 89 (1976).
2. D. Psaltis, "Incoherent Electrooptic Image Correlator," *Opt. Eng.* 23, 12 (1984).
3. N. V. Kukhtarev, V. B. Markov, S. G. Odulov, and M. S. Soskin, "Holographic Storage in Crystals. I: Steady State," *Ferroelectrics* 22, 949 (1979).
4. J. Feinberg, D. Heiman, A. R. Tanguay, and R. W. Hellwarth, "Photo Refractive Effects and Light-Induced Charge Migration in Barium Titanate," *J. Appl. Phys.* 51, 1297 (1980).
5. M. Cronin-Golomb, "Large Nonlinearities in Four-Wave Mixing in Photorefractive Crystals and Applications in Passive Optical Phase Conjugation," Ph.D. Thesis, California Institute of Technology (Mar. 1983).

Optical implementation of the Hopfield model

Nabil H. Farhat, Demetri Psaltis, Aluizio Prata, and Eung Paek

Optical implementation of content addressable associative memory based on the Hopfield model for neural networks and on the addition of nonlinear iterative feedback to a vector-matrix multiplier is described. Numerical and experimental results presented show that the approach is capable of introducing accuracy and robustness to optical processing while maintaining the traditional advantages of optics, namely, parallelism and massive interconnection capability. Moreover a potentially useful link between neural processing and optics that can be of interest in pattern recognition and machine vision is established.

I. Introduction

It is well known that neural networks in the eye-brain system process information in parallel with the aid of large numbers of simple interconnected processing elements, the neurons. It is also known that the system is very adept at recognition and recall from partial information and has remarkable error correction capabilities.

Recently Hopfield described a simple model¹ for the operation of neural networks. The action of individual neurons is modeled as a thresholding operation and information is stored in the interconnections among the neurons. Computation is performed by setting the state (on or off) of some of the neurons according to an external stimulus and, with the interconnections set according to the recipe that Hopfield prescribed, the state of all neurons that are interconnected to those that are externally stimulated spontaneously converges to the stored pattern that is most similar to the external input. The basic operation performed is a nearest-neighbor search, a fundamental operation for pattern recognition, associative memory, and error correction. A remarkable property of the model is that powerful global computation is performed with very simple, identical logic elements (the neurons). The interconnections provide the computation power to these simple logic elements and also enhance dramatically the stor-

age capacity; approximately $N/4 \ln N$ bits/neuron can be stored in a network in which each neuron is connected to N others.² Another important feature is that synchronization among the parallel computing elements is not required, making concurrent, distributed processing feasible in a massively parallel structure. Finally, the model is insensitive to local imperfections such as variations in the threshold level of individual neurons or the weights of the interconnections.

Given these characteristics we were motivated to investigate the feasibility of implementing optical information processing and storage systems that are based on this and other similar models of associative memory.^{3,4} Optical techniques offer an effective means for the implementation of programmable global interconnections of very large numbers of identical parallel logic elements. In addition, emerging optical technologies such as 2-D spatial light modulators, optical bistability, and thin-film optical amplifiers appear to be very well suited for performing the thresholding operation that is necessary for the implementation of the model.

The principle of the Hopfield model and its implications in optical information processing have been discussed earlier.^{5,6} Here we review briefly the main features of the model, give as an example the results of a numerical simulation, describe schemes for its optical implementation, then present experimental results obtained with one of the schemes and discuss their implications as a content addressable associative memory (CAM).

II. Hopfield Model

Given a set of M bipolar, binary $(1, -1)$ vectors $\mathbf{v}_i^{(m)}$, $i = 1, 2, 3 \dots N$, $m = 1, 2, 3 \dots M$, these are stored in a synaptic matrix in accordance with the recipe

$$T_{ij} = \sum_{m=1}^M v_i^{(m)} v_j^{(m)}, \quad i, j = 1, 2, 3 \dots N, \quad T_{ii} = 0, \quad (1)$$

$\mathbf{v}_i^{(m)}$ are referred to as the nominal state vectors of the

Nabil Farhat is with University of Pennsylvania, Moore School of Electrical Engineering, Philadelphia, Pennsylvania 19104; the other authors are with California Institute of Technology, Electrical Engineering Department, Pasadena, California 91125.

Received 24 December 1984.

0003-6935/85/101469-07\$02.00/0.

© 1985 Optical Society of America.

memory. If the memory is addressed by multiplying the matrix T_{ij} with one of the state vectors, say $v_i^{(mo)}$, it yields the estimate

$$\phi_i^{(mo)} = \sum_j T_{ij} v_j^{(mo)} \quad (2)$$

$$\begin{aligned} &= \sum_{j \neq i} \sum_m v_i^{(m)} v_j^{(m)} v_j^{(mo)} \\ &= (N-1)v_i^{(mo)} + \sum_{m \neq m_0} \alpha_{m,m_0} v_i^{(m)}, \end{aligned} \quad (3)$$

where

$$\alpha_{m,m_0} = \sum_j v_j^{(mo)} v_j^{(m)}.$$

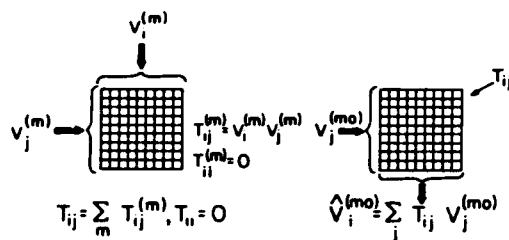
$\phi_i^{(mo)}$ consists of the sum of two terms: the first is the input vector amplified by $(N-1)$; the second is a linear combination of the remaining stored vectors and it represents an unwanted cross-talk term. The value of the coefficients α_{m,m_0} is equal to $\sqrt{N-1}$ on the average (the standard deviation of the sum of $N-1$ random bits), and since $(M-1)$ such coefficients are randomly added, the value of the second term will on the average be equal to $\sqrt{(M-1)(N-1)}$. If N is sufficiently larger than M , with high probability the elements of the vector $\phi_i^{(mo)}$ will be positive if the corresponding elements of $v_i^{(mo)}$ are equal to $+1$ and negative otherwise. Thresholding of $\phi_i^{(mo)}$ will therefore yield $v_i^{(mo)}$:

$$v_i^{(mo)} = \text{sgn}[\phi_i^{(mo)}] = \begin{cases} +1 & \text{if } \phi_i^{(mo)} > 0 \\ -1 & \text{otherwise.} \end{cases} \quad (4)$$

When the memory is addressed with a binary valued vector that is not one of the stored words, the vector-matrix multiplication and thresholding operation yield an output binary valued vector which, in general, is an approximation of the stored word that is at the shortest Hamming distance from the input vector. If this output vector is fed back and used as the input to the memory, the new output is generally a more accurate version of the stored word and continued iteration converges to the correct vector.

The insertion and readout of memories described above are depicted schematically in Fig. 1. Note that in Fig. 1(b) the estimate $\phi_i^{(mo)}$ can be viewed as the weighted projection of T_{ij} . Recognition of an input vector that corresponds to one of the state vectors of the memory or is close to it (in the Hamming sense) is manifested by a stable state of the system. In practice unipolar binary (0,1) vectors or words $b_i^{(m)}$ of bit length N may be of interest. The above equations are then applicable with $[2b_i^{(m)} - 1]$ replacing $v_i^{(m)}$ in Eq. (1) and $b_i^{(mo)}$ replacing $v_i^{(mo)}$ in Eq. (2). For such vectors the SNR of the estimate $\phi_i^{(mo)}$ can be shown to be lower by a factor of $\sqrt{2}$.¹

An example of the T_{ij} matrix formed from four binary unipolar vectors, each being $N = 20$ bits long, is given in Fig. 2 along with the result of a numerical simulation of the process of initializing the memory matrix with a partial version of $b_i^{(4)}$ in which the first eight digits of $b_i^{(4)}$ are retained and the remainder set to zero. The Hamming distance between the initializing vector and $b_i^{(4)}$ is 6 bits and it is 9 or more bits for the other three



(a) READ-IN (b) READ-OUT
Fig. 1. (a) Insertion and (b) readout of memories.

stored vectors. It is seen that the partial input is recognized as $b_i^{(4)}$ in the third iteration and the output remains stable as $b_i^{(4)}$ thereafter. This convergence to a stable state generally persists even when the T_{ij} matrix is binarized or clipped by replacing negative elements by minus ones and positive elements by plus ones evidencing the robustness of the CAM. A binary synaptic matrix has the practical advantage of being more readily implementable with fast programmable spatial light modulators (SLM) with storage capability such as the Litton Lightmod.⁷ Such a binary matrix, implemented photographically, is utilized in the optical implementation described in Sec. III and evaluated in Sec. IV of this paper.

Several schemes for optical implementation of a CAM based on the Hopfield model have been described earlier.⁵ In one of the implementations an array of light emitting diodes (LEDs) is used to represent the logic elements or neurons of the network. Their state (on or off) can represent unipolar binary vectors such as the state vectors $b_i^{(m)}$ that are stored in the memory matrix T_{ij} . Global interconnection of the elements is realized as shown in Fig. 3(a) through the addition of nonlinear feedback (thresholding, gain, and feedback) to a conventional optical vector-matrix multiplier⁸ in which the array of LEDs represents the input vector and an array of photodiodes (PDs) is used to detect the output vector. The output is thresholded and fed back in parallel to drive the corresponding elements of the LED array. Multiplication of the input vector by the T_{ij} matrix is achieved by horizontal imaging and vertical smearing of the input vector that is displayed by the LEDs on the plane of the T_{ij} mask [by means of an anamorphic lens system omitted from Fig. 3(a) for simplicity]. A second anamorphic lens system (also not shown) is used to collect the light emerging from each row of the T_{ij} mask on individual photosites of the PD array. A bipolar T_{ij} matrix is realized in incoherent light by dividing each row of the T_{ij} matrix into two subrows, one for positive and one for negative values and bringing the light emerging from each subrow to focus on two adjacent photosites of the PD array that are electrically connected in opposition as depicted in Fig. 3(b). In the system shown in Fig. 3(a), feedback is achieved by electronic wiring. It is possible and preferable to dispose of electronic wiring altogether and replace it by optical feedback. This can be achieved by combining the PD and LED arrays in a single compact hybrid or

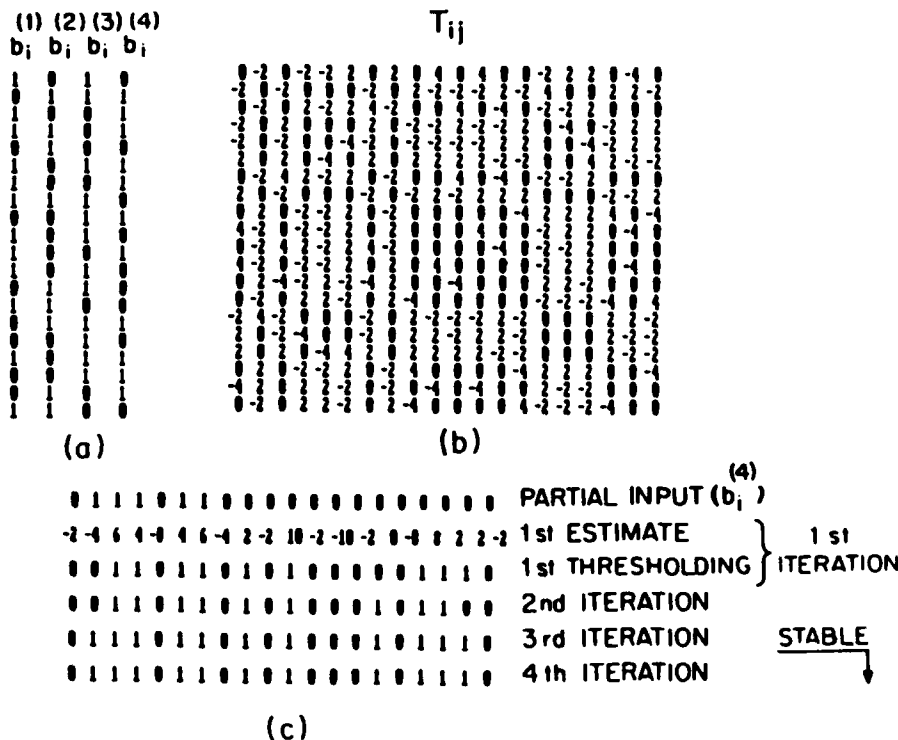


Fig. 2. Numerical example of recovery from partial input; $N = 20$, $M = 4$. (a) Stored vectors, (b) memory or (synaptic) matrix, (c) results of initializing with a partial version of $b_i^{(4)}$.

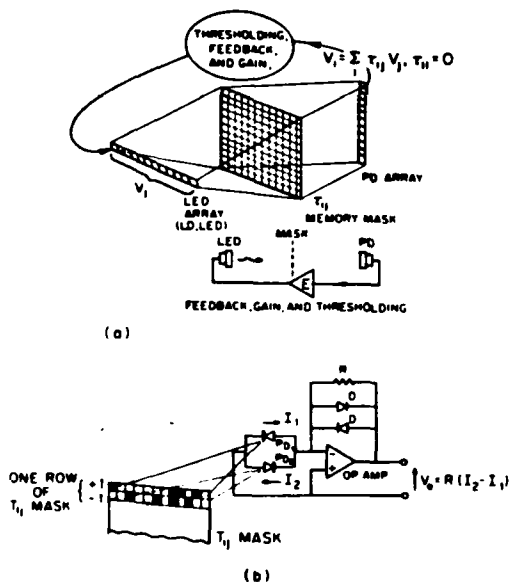


Fig. 3. Concept for optical implementation of a content addressable memory based on the Hopfield model. (a) Matrix-vector multiplier incorporating nonlinear electronic feedback. (b) Scheme for realizing a binary bipolar memory mask transmittance in incoherent light.

monolithic structure that can also be made to contain all ICs for thresholding, amplification, and driving of LEDs. Optical feedback becomes even more attractive when we consider that arrays of nonlinear optical light amplifiers with internal feedback⁹ or optical bistability

devices (OBDs)¹⁰ can be used to replace the PD/LED arrays. This can lead to simple compact CAM structures that may be interconnected to perform higher-order computations than the nearest-neighbor search performed by a single CAM.

We have assembled a simple optical system that is a variation of the scheme presented in Fig. 3(a) to simulate a network of $N = 32$ neurons. The system, details of which are given in Figs. 5-8, was constructed with an array of thirty-two LEDs and two multichannel silicon PD arrays, each consisting of thirty-two elements. Twice as many PD elements as LEDs are needed in order to implement a bipolar memory mask transmittance in incoherent light in accordance with the scheme of Fig. 3(b). A bipolar binary T_{ij} mask was prepared for $M = 3$ binary state vectors. The three vectors or words chosen, their Hamming distances from each other, and the resulting T_{ij} memory matrix are shown in Fig. 4. The mean Hamming distance between the three vectors is 16. A binary photographic transparency of 32×64 square pixels was computer generated from the T_{ij} matrix by assigning the positive values in any given row of T_{ij} to transparent pixels in one subrow of the mask and the negative values to transparent pixels in the adjacent subrow. To insure that the image of the input LED array is uniformly smeared over the memory mask it was found convenient to split the mask in two halves, as shown in Fig. 5, and to use the resulting submasks in two identical optical arms as shown in Fig. 6. The size of the subrows of the memory submasks was made exactly equal to the element size of the PD arrays in the vertical direction which were placed in register

Stored words:

Word 1 : 1 1 1 0 0 0 0 1 0 1 0 1 1 1 0 1 1 1 0 1 1 0 0 0 0 0 1 0
 Word 2 : 0 1 1 0 0 0 0 0 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 0 1 1 0 0
 Word 3 : 1 0 1 1 0 0 1 1 1 1 1 1 1 1 0 0 0 1 0 1 1 0 0 0 0 1 1 0 0 0

Hamming distance from word to word:

WORD	1	2	3
1		0	15
2	15		0
3	14	19	

Clipped memory matrix:

0 -1 1 1 -1 -1 1 1 1 1 -1 1 1 1 1 -1 -1 1 1 1 -1 -1 1 -1 -1 -1 -1 -1
 -1 0 1 -1 -1 -1 -1 -1 -1 -1 -1 -1 1 1 1 -1 1 1 1 1 1 -1 -1 1 -1 1 -1
 1 1 0 -1 -1 -1 -1 -1 1 1 1 1 1 -1 1 -1 1 -1 1 -1 1 -1 -1 -1 1 -1 1 -1
 1 -1 -1 0 1 1 1 1 1 1 1 1 -1 -1 -1 -1 -1 -1 -1 -1 1 1 -1 1 -1 1 -1
 -1 -1 -1 1 0 1 1 -1 -1 -1 -1 -1 -1 1 -1 1 -1 -1 -1 1 1 1 -1 1 -1 1 -1
 -1 -1 1 1 1 0 1 -1 -1 -1 -1 -1 -1 1 -1 1 -1 -1 -1 1 1 -1 1 1 -1 1 -1
 1 -1 1 1 1 1 0 1 1 1 1 1 -1 -1 -1 -1 -1 -1 -1 -1 1 1 -1 1 -1 1 -1
 1 -1 1 1 -1 -1 1 0 1 1 1 1 1 -1 -1 -1 1 1 1 -1 -1 1 -1 -1 -1 -1 -1
 1 -1 -1 1 1 1 1 1 0 1 1 1 1 -1 -1 -1 1 -1 -1 -1 -1 1 1 -1 1 -1 1 -1
 1 -1 1 1 -1 -1 1 1 1 1 1 -1 0 1 1 1 -1 -1 1 1 1 -1 -1 1 -1 -1 -1 -1
 1 -1 1 1 -1 -1 1 1 1 1 -1 0 1 1 1 -1 -1 1 1 1 -1 -1 1 -1 -1 -1 -1
 1 1 1 -1 -1 -1 1 1 1 1 1 1 0 -1 1 -1 -1 1 -1 1 -1 -1 -1 1 -1 1 -1
 1 -1 -1 1 1 1 1 1 1 1 1 -1 0 -1 -1 -1 -1 -1 -1 1 1 -1 1 1 -1 1 -1
 -1 1 1 -1 -1 -1 -1 -1 -1 -1 -1 1 -1 0 1 1 1 1 1 1 -1 1 -1 1 -1 1 -1
 1 1 -1 1 1 1 -1 1 -1 -1 -1 1 1 -1 1 0 -1 1 1 -1 -1 1 -1 -1 1 1 1
 1 -1 1 1 -1 -1 1 1 1 1 -1 1 1 1 -1 0 1 1 -1 -1 1 -1 -1 -1 -1 -1
 1 1 -1 1 1 1 -1 1 -1 -1 1 1 -1 1 1 -1 -1 1 1 -1 -1 -1 1 1 1
 1 1 1 -1 -1 -1 1 1 1 1 1 1 -1 1 -1 1 1 -1 -1 -1 1 -1 -1 1 1 -1
 1 1 1 -1 -1 -1 -1 1 1 1 1 1 -1 1 -1 1 1 0 -1 1 -1 -1 -1 1 -1
 -1 1 -1 -1 1 1 -1 -1 -1 -1 -1 -1 -1 -1 -1 0 1 -1 1 1 1 1 1 1
 1 1 1 -1 -1 -1 -1 -1 -1 -1 -1 1 1 1 1 1 1 0 1 1 -1 -1 1 -1 -1
 1 1 -1 1 1 1 -1 1 -1 1 1 1 -1 1 1 -1 -1 1 0 -1 -1 -1 1 1 1
 -1 1 -1 1 1 1 -1 -1 -1 1 1 -1 -1 1 1 -1 -1 1 1 -1 0 1 1 1 1
 -1 1 1 1 -1 1 -1 1 -1 1 1 -1 -1 1 1 -1 -1 1 1 0 1 -1 -1 -1
 -1 1 -1 1 1 1 -1 -1 -1 1 1 -1 -1 1 1 -1 -1 1 1 -1 1 0 1 1 1
 -1 1 1 1 1 1 -1 -1 -1 -1 -1 -1 1 1 -1 1 1 -1 1 1 1 0 -1 1
 -1 1 -1 1 1 1 1 1 -1 -1 -1 1 1 1 1 1 1 1 1 1 -1 1 -1 0 -1
 -1 -1 1 1 1 1 1 1 -1 -1 -1 -1 1 1 1 1 1 1 1 1 1 1 -1 1 -1 0

Fig. 4. Stored words, their Hamming distances, and their clipped T_{ij} memory matrix.

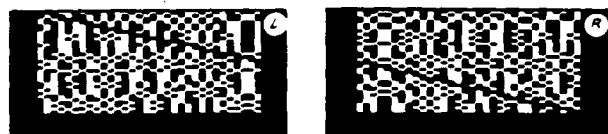


Fig. 5. Two halves of T_{ij} memory mask.

against the masks. Light emerging from each subrow of a memory submask was collected (spatially integrated) by one of the vertically oriented elements of the multichannel PD array. In this fashion the anamorphic optics required in the output part of Fig. 3(a) are disposed of, resulting in a more simple and compact system. Pictorial views of the input LED array and the

two submask/PD array assemblies are shown in Figs. 7(a) and (b), respectively. In Fig. 7(b) the left memory submask/PD array assembly is shown with the submask removed to reveal the silicon PD array situated behind it. All electronic circuits (amplifiers, thresholding comparators, LED drivers, etc.) in the thirty-two parallel feedback channels are contained in the electronic amplification and thresholding box shown in Fig. 6(a) and in the boxes on which the LED array, and the two submask/PD array assemblies are mounted (see Fig. 7). A pictorial view of a composing and display box is shown in Fig. 8. This contains an arrangement of thirty-two switches and a thirty-two element LED display panel whose elements are connected in parallel to the input LED array. The function of this box is to compose and

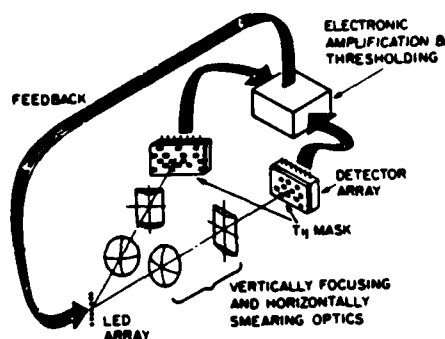


Fig. 6. Arrangement for optical implementation of the Hopfield model: (a) optoelectronic circuit diagram, (b) pictorial view.

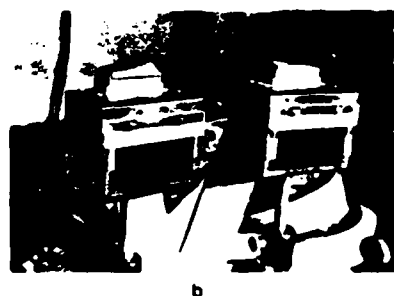
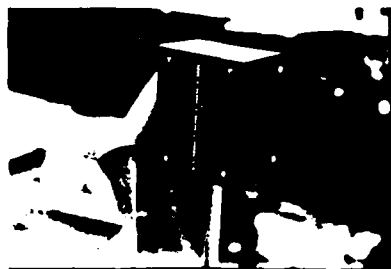


Fig. 7. Views of (a) input LED array and (b) memory submask/PD array assemblies.

display the binary input word or vector that appears on the input LED array of the system shown in Fig. 7(a). Once an input vector is selected it appears displayed on the composing box and on the input LED box simultaneously. A single switch is then thrown to release the system into operation with the composed vector as the

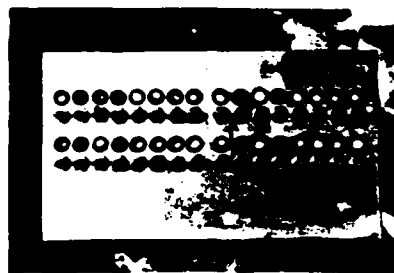


Fig. 8. Word composer and display box.

initializing vector. The final state of the system, the output, appears after a few iterations displayed on the input LED array and the display box simultaneously. The above procedure provides for convenient exercising of the system in order to study its response vs stimulus behavior. An input vector is composed and its Hamming distance from each of the nominal state vectors stored in the memory is noted. The vector is then used to initialize the CAM as described above and the output vector representing the final state of the CAM appearing, almost immediately, on the display box is noted. The response time of the electronic feedback channels as determined by the 3-dB roll-off of the amplifiers was ~ 60 msec. Speed of operation was not an issue in this study, and thus low response time was chosen to facilitate the experiment.

IV. Results

The results of exercising and evaluating the performance of the system we described in the preceding section are tabulated in Table I. The first run of initializing vectors used in exercising the system were error laden versions of the first word $b_i^{(1)}$. These were obtained from $b_i^{(1)}$ by successively altering (switching) the states of 1, 2, 3 . . . up to N of its digits starting from the N th digit. In doing so the Hamming distance between the initializing vector and $b_i^{(1)}$ is increased linearly in unit steps as shown in the first column of Table I whereas, on the average, the Hamming distance between all these initializing vectors and the other two state vectors remained approximately the same, about $N/2 = 16$. The final states of the memory, i.e., the steady-state vectors displayed at the output of the system (the composing and display box) when the memory is prompted by the initializing vectors, are listed in column 2 of Table I. When the Hamming distance of the initializing vector from $b_i^{(1)}$ is < 11 , the input is always recognized correctly as $b_i^{(1)}$. The CAM is able therefore to recognize the input vector as $b_i^{(1)}$ even when up to 11 of its digits (37.5%) are wrong. This performance is identical to the results obtained with a digital simulation shown in parenthesis in column 2 for comparison. When the Hamming distance is increased further to values lying between 12 and 22, the CAM is confused and identifies erroneously other state vectors, mostly $b_i^{(3)}$, as the input. In this range, the Hamming distance of the initializing vectors from any of the stored vectors is approximately equal making it more difficult for the CAM to decide. Note that the performance of

Table I. Optical CAM Performance

Hamming distance of initializing vector from $b_i^{(m)}$	Recognized vector ($m = 1$)	Recognized vector ($m = 2$)	Recognized vector ($m = 3$)
0	1 (1)	2 (2)	3 (3)
1	1 (1)	2 (2)	3 (3)
2	1 (1)	2 (2)	3 (3)
3	1 (1)	2 (2)	3 (3)
4	1 (1)	2 (2)	3 (3)
5	1 (1)	2 (2)	3 (3)
6	1 (1)	2 (2)	3 (3)
7	1 (1)	2 (2)	3 (3)
8	1 (1)	2 (2)	3 (3)
9	1 (1)	2 (2)	3 (3)
10	1 (1)	1 (1)	3 (3)
11	1 (1)	2 (2)	3 (3)
12	3 (3)	3 (3)	3 (3)
13	3 (3)	3 (3)	3 (2)
14	3 (3)	1 (1)	3 (2)
15	1 (OSC)	1 (1)	2 (2)
16	3 (OSC)	1 (1)	2 (2)
17	3 (OSC)	1 (OSC)	2 (2)
18	3 (3)	1 (2)	3 (OSC)
19	3 (2)	2 (2)	2 (2)
20	3 (1)	2 (2)	2 (OSC)
21	1, 2 (1)	2 (2)	3 (OSC)
22	3 (1)	2 (2)	3 (OSC)
23	1 (1)	2 (2)	3 (OSC)
24	1 (1)	2 (2)	3 (3)
25	1 (1)	2 (2)	3 (3)
26	1 (1)	2 (2)	3 (3)
27	1 (1)	2 (2)	3 (3)
28	1 (1)	2 (2)	3 (3)
29	1 (1)	2 (2)	3 (3)
30	1 (1)	2 (2)	3 (3)
31	1 (1)	2 (2)	3 (3)
32	1 (1)	2 (2)	3 (3)

the CAM and results of digital simulation in this range of Hamming distance are comparable except for the appearance of oscillations (designated by OSC) in the digital simulation when the outcome oscillated between several vectors that were not the nominal state vectors of the CAM. Beyond a Hamming distance of 22 both the optical system and the digital simulation identified the initializing vectors as the complement $\bar{b}_i^{(1)}$ of $b_i^{(1)}$. This is expected because it can be shown using Eq. (1) that the T_{ij} matrix formed from a set of vectors $b_i^{(m)}$ is identical to that formed by the complementary set $\bar{b}_i^{(m)}$. The complementary vector can be viewed as a contrast reversed version of the original vector in which zeros and ones are interchanged. Recognition of a complementary state vector by the CAM is analogous to our recognizing a photographic image from the negative.

Similar results of initializing the CAM with error laden versions of $b_i^{(2)}$ and $b_i^{(3)}$ were also obtained. These are presented in columns 2 and 3 of Table I. Here again we see when the Hamming distance of the initializing vector from $b_i^{(3)}$, for example, ranged between 1 and 14, the CAM recognized the input correctly as $b_i^{(3)}$ as shown in column 3 of the table and as such it did slightly better than the results of digital simulation. Oscillatory behavior is also observed here in the digital simulation when the range of Hamming distance between the ini-

tializing vector from all stored vectors approached the mean Hamming distance between the stored vectors. Beyond this range the memory recognizes the input as the complementary of $b_i^{(3)}$.

In studying the results presented in Table I several observations can be made: The optically implemented CAM is working as accurately as the digital simulations and perhaps better if we consider the absence of oscillations. These are believed to be suppressed in the system because of the nonsharp thresholding performed by the smoothly varying nonlinear transfer function of electronic circuits compared with the sharp thresholding in digital computations. The smooth nonlinear transfer function and the finite time constant of the optical system provide a relaxation mechanism that substitutes for the role of asynchronous switching required by the Hopfield model. Generally the system was able to conduct successful nearest-neighbor search when the inputs to the system are versions of the nominal state vectors containing up to ~30% error in their digits. It is worth noting that this performance is achieved in a system built from off-the-shelf electronic and optical components and with relatively little effort in optimizing and fine tuning the system for improved accuracy, thereby confirming the fact that accurate global computation can be performed with relatively inaccurate individual components.

V. Discussion

The number M of state vectors of length N that can be stored at any time in the interconnection matrix T_{ij} is limited to a fraction of N . An estimate of $M \approx 0.1N$ is indicated in simulations involving a hundred neurons or less¹ and a theoretical estimate of $M \approx N/4 \ln N$ has recently been obtained.² It is worthwhile to consider the number of bits that can be stored per interconnection or per neuron. The number of pixels required to form the interconnection matrix is N^2 . Since such a T_{ij} memory matrix can store up to $M \approx N/4 \ln N$ (N -tuples), the number of bits stored is $MN = N^2/4 \ln N$. The number of bits stored per memory matrix element or interconnection is $MN/N^2 = (4 \ln N)^{-1}$, while the number of bits stored per neuron is $MN/N = M$.

The number of stored memories that can be searched for a given initializing input can be increased by using a dynamic memory mask that is rapidly addressed with different T_{ij} matrices each corresponding to different sets of M vectors. The advantage of programmable SLMs for realizing this goal are evident. For example, the Litton Lightmod (magneto-optic light modulator), which has nonvolatile storage capability and can provide high frame rates, could be used. A frame rate of 60 Hz is presently specified for commercially available units of 128×128 pixels which are serially addressed.⁷ Units with 256×256 pixels are also likely to be available in the near future with the same frame rate capability. Assuming a memory mask is realized with a Litton Lightmod of 256×256 pixels we have $N = 256$, $M \approx 0.1N \approx 26$ and a total of $26 \times 60 = 1560$ vectors can be searched or compared per second against an initializing input vector. Speeding up the frame rate of the Litton

Lightmod to increase memory throughput beyond the above value by implementing parallel addressing schemes is also possible. Calculations show that the maximum frame rate possible for the device operating in reflection mode with its drive lines heat sunk is 10 kHz.⁷ This means the memory throughput estimated above can be increased to search 2.6×10^5 vectors/sec, each being 256 bits long, or a total of 6.7×10^8 bits/sec. This is certainly a respectable figure, specially when we consider the error correcting capability and the associative addressing mode of the Hopfield model; i.e., useful computation is performed in addition to memory addressing.

The findings presented here show that the Hopfield model for neural networks and other similar models for content addressable and associative memory fit well the attributes of optics, namely, parallel processing and massive interconnection capabilities. These capabilities allow optical implementation of large neural networks based on the model. The availability of nonlinear or bistable optical light amplifiers with internal feedback, optical bistability devices, and nonvolatile high speed spatial light modulators could greatly simplify the construction of optical CAMs and result in compact modules that can be readily interconnected to perform more general computation than nearest-neighbor search. Such systems can find use in future generation computers, artificial intelligence, and machine vision.

The work described in this paper was performed while one of the authors, N.F., was on scholarly leave at the California Institute of Technology. This author wishes to express his appreciation to CIT and the University of Pennsylvania for facilitating his sabbatical leave. The work was supported in part by the Army Research Office and in part by the Air Force Office of Scientific Research.

The subject matter of this paper is based on a paper presented at the OSA Annual Meeting, San Diego, Oct. 1984.

References

1. J. J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," *Proc. Natl. Acad. Sci. USA* **79**, 2554 (1982).
2. R. J. McEliece, E. C. Posner, and S. Venkatesh, California Institute of Technology, Electrical Engineering Department; private communication.
3. G. E. Hinton and J. A. Anderson, *Parallel Models of Associative Memory* (LEA Publishers, Hillsdale, N.J., 1981).
4. T. Kohonen, *Content Addressable Memories* (Springer, New York, 1980).
5. D. Psaltis and N. Farhat, "A New Approach to Optical Information Processing Based On the Hopfield Model," in *Technical Digest, ICO-13 Conference, Sapporo* (1984), p. 24.
6. D. Psaltis and N. Farhat, "Optical Information Processing Based on an Associative-Memory Model of Neural Nets with Thresholding and Feedback," *Opt. Lett.* **10**, 98 (1985).
7. W. Ross, D. Psaltis, and R. Anderson, "Two-Dimensional Magneto-Optic Spatial Light Modulator For Signal Processing," *Opt. Eng.* **22**, 485 (1983).
8. J. W. Goodman, A. R. Dias, and L. M. Woody, "Fully Parallel, High-Speed Incoherent Optical Method for Performing Discrete Fourier Transforms," *Opt. Lett.* **2**, 1 (1978).
9. Z. Porada, "Thin Film Light Amplifier with Optical Feedback," *Thin Solid Films* **109**, 213 (1983).
10. H. M. Gibbs *et al.*, "Optical Bistable Devices: The Basic Components of All-Optical Circuits," *Proc. Soc. Photo-Opt. Instrum. Eng.* **269**, 75 (1981).

END

12-86

DTIC